The PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

VOL. 46 SEVENTH SERIES No. 382 November 1955

A Journal of Theoretical Experimental and Applied Physics

PRICE 15s. 0d.
Annual Subscription £8 0s. 0d. payable in advance

NTED AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET ST., LONDON, E.C.4.

CXXVIII. *The Remanent Magnetization of Haematite Powders*

By E. P. Wohlfarth
Department of Mathematics, Imperial College, London*

[Received July 21, 1955]

Summary

Calculations are described of the remanent magnetization and the remanence after demagnetization as a function of field strength for powders of haematite, αFe_2O_3. The calculations are based on the assumptions that the powder particles are single domains and that their hysteretic properties are due to magnetocrystalline anisotropy with threefold symmetry in the basal crystal plane. The calculations are intended for eventual comparison with experiments now in progress.

§ 1. Introduction

The oxide mineral haematite, αFe_2O_3, is one of the most important ferromagnetic constituents of rocks. Recently experiments have been started in the Department of Physics of this College, by Professor P. M. S. Blackett, F.R.S. and Mr. G. Haigh, designed to investigate more fully the magnetic properties of pure haematite powders. In particular, preliminary experiments have been carried out and more systematic ones are planned on the variation with field strength of the remanent magnetization and of the remanence after demagnetization, both at and above room temperature and also at temperatures below the Morin transition point (250°–260°K, Morin 1950). The present paper is intended to supply theoretical values for eventual comparison with the experimental results. The calculations are based on a number of simplifying assumptions which it may be necessary to modify in the light of these results.

The following are the relevant properties of haematite, the assumptions used in the present calculations being briefly indicated and discussed:

(1) The crystal structure of haematite is rhombohedral, with an axis of threefold symmetry perpendicular to a basal symmetry plane; the arrangement of iron and oxygen ions in the lattice is described by Nicholls (1955).

(2) The magnetization σ depends on temperature and field strength; Néel and Pauthenet (1952, cf. also Néel 1953) have shown that in moderately high fields the relation for σ is

$$\sigma(T, H)=\sigma_0(T)+\chi(T)H. \quad . \quad . \quad . \quad . \quad (1.1)$$

* Communicated by the Author.

The mass susceptibility $\chi(T)$ is anisotropic below the Morin transition point, 250°–260°K, being approximately zero along the trigonal axis and isotropic and approximately constant, about 20×10^{-6}, in the basal plane. Above this temperature $\chi(T)$ is constant and equal to about 20×10^{-6}, both along the trigonal axis and in the basal plane, up to high temperatures including the Curie point, about 950°K. These properties indicate that the paramagnetism of haematite is actually an antiferromagnetism (Néel 1953), the antiferromagnetic axis being parallel to the trigonal axis below the Morin transition and perpendicular to it above. This result is also borne out by neutron diffraction studies (Shull *et al.* 1951).

The magnetization $\sigma_0(T)$ may be written as the sum of two parts,

$$\sigma_0=\sigma_i+\sigma_a. \qquad (1.2)$$

Here σ_i is an isotropic component having a saturation value about 0·2 e.m.u. per unit mass (1·0 per unit volume) and retaining this value both along the trigonal axis and perpendicular to it over a wide temperature range, finally vanishing at the Curie point. On the other hand σ_a is zero below the Morin transition point but rises to a value about 0·3 e.m.u. per unit mass (1·6 per unit volume) at this temperature, is anisotropic in the basal plane and finally vanishes at the Curie point. Along the trigonal axis no component of magnetization corresponding to σ_a can be observed in moderate fields at any temperature. The ferromagnetism of haematite is closely linked to its antiferromagnetism, but its origin is still uncertain. Néel (1949, 1953, 1955) ascribes it to one of two principal causes, ferromagnetic impurities such as Fe_3O_4 or γFe_2O_3, or defects in the lattice planes containing the iron ions.

The hysteretic properties of haematite above the Morin transition point and in moderate fields may be assumed to arise from the anisotropic component of magnetization σ_a in the basal plane. Since no information about the nature of the easy axes of magnetization nor of the magnitude of the corresponding anisotropy coefficients in this plane is available, it is here assumed on grounds of symmetry that there are three such easy axes making equal angles with each other. The simplest expression for the anisotropy energy commensurate with the symmetry is assumed (see relation (2.2) below) and the anisotropy coefficient K thus introduced is taken to be an arbitrary parameter the value of which should eventually be obtained by experiment (cf. § 2).

Measurements of the magnetization in any field, e.g. the remanent magnetization, give the contributions due to both the anisotropic and the isotropic component. In order to obtain the contribution due to the anisotropic component alone a very convenient method of analysing the experimental results is available, suggested by Mr. G. Haigh: Since the isotropic component of magnetization σ_i varies little with temperature near the Morin transition point, while the anisotropic component σ_a is zero below and approximately constant near and above this temperature,

measurements of the magnetization on either side of the transition temperature should yield at once the separate contributions σ_i and σ_a. The isotropic component is here assumed to make no contribution to the hysteretic properties, i.e. it is assumed to be reversible; the preliminary experiments indicate that this assumption is not strictly true.

Although the trigonal axis of haematite is a very difficult axis of magnetization, so that in moderate fields the magnetization vector of a crystal remains parallel to the basal plane, application of high fields inclined to the basal plane must cause some rotations of the magnetization vector towards the trigonal axis. Estimates of the anisotropy field corresponding to magnetization along the trigonal axis are provided by the results of ferromagnetic resonance measurements due to Anderson *et al.* (1954) who derive the value 3×10^4 oersted. The present experiments have been carried out in much more moderate fields so that it may be assumed as an approximation that the magnetization vector is always strictly parallel to the basal plane.

To sum up, the hysteretic properties are here assumed to arise solely from the anisotropic component of magnetization in the basal plane, the isotropic component in that plane being taken to be infinitely ' soft ' and the anisotropic component parallel to the trigonal axis infinitely ' hard '.

(3) Although the mean size of the powder particles used in the experiments has not yet been determined, it may be assumed that they are single domains so that the magnetization change is one of rotation. No accurate estimate of the critical size of haematite powders has yet been made since the value of the anisotropy coefficient K in the basal plane is not known, but the critical size, which is proportional to $K^{1/2}/I_0$, is probably large (a very rough estimate on the basis of the approximate K value deduced in § 2 is 1 μ) owing to the small value of I_0, the saturation magnetization of the anisotropic component, 1·6 gauss. If the ferromagnetism of haematite is due to impurities or lattice defects it may be that each powder particle consists of smaller ferromagnetic islands embedded in a non-magnetic matrix (Néel 1955). These islands are then even more likely to be single domains.

(4) Since the hysteretic properties of haematite are controlled by magnetocrystalline anisotropy, the saturation magnetization being very small, it is justifiable to neglect the effect of particle interaction on these properties (cf. Wohlfarth 1955 a).

(5) In an assembly of powder particles it may be assumed that the orientations both of the trigonal axes and of the easy axes of magnetization in the basal planes are random.

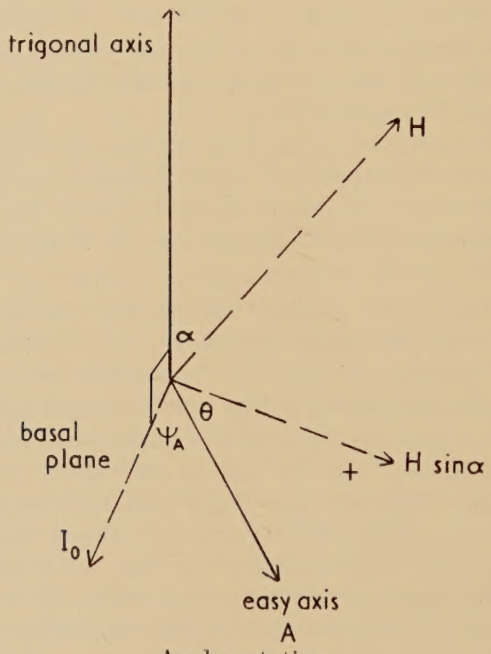
In § 2 calculations are given of the curve relating the remanent magnetization I_R with a reduced field strength and in § 3 of the curve relating the remanence after demagnetization I_D with field strength.

§ 2. Remanent Magnetization

In the experiments a field of strength H is applied to the initially demagnetized powder and the remanent magnetization I_R measured after removal of the field. The curve obtained below gives the contribution to the remanence of the anisotropic component of magnetization in the basal plane, obtainable from the measured values as discussed in § 1.

Figure 1 shows the angle notation, α being the angle between **H** and the trigonal axis, $\theta(0 \leqslant \theta \leqslant \pi/3)$ that between the positive direction of the component of **H** in the basal plane and the nearest easy axis A and ψ_A that between the magnetization vector $\mathbf{I}_0$ and A.

Fig. 1



Angle notation.

The total energy is the sum of the energy in the external field E_H and the anisotropy energy E_A, where

$$E_H = -HI_0 \sin\alpha \cos(\theta+\psi_A). \quad \ldots\ldots \quad (2.1)$$

Also the lowest non-vanishing term in the general expression for the anisotropy energy, which alone is considered here, is of the form

$$E_A = K \sin^2(3\psi_A/2), \quad \ldots\ldots \quad (2.2)$$

K being the anisotropy constant. For equilibrium

$$h \sin\alpha \sin(\theta+\psi_A) + \tfrac{1}{2} \sin 3\psi_A = 0, \quad \ldots\ldots \quad (2.3)$$

$$h \sin\alpha \cos(\theta+\psi_A) + \tfrac{3}{2} \cos 3\psi_A > 0, \quad \ldots\ldots \quad (2.4)$$

where

$$h = HI_0/3K. \quad \ldots\ldots \quad (2.5)$$

The rotations of the magnetization vector with changing field strength become discontinuous when the inequality in (2.4) becomes an equality, the critical value of ψ_A, denoted by $\psi_0(\theta)$, being given by the relation

$$3\tan(\theta+\psi_0)=\tan 3\psi_0. \qquad \ldots\ldots \quad (2.6)$$

The corresponding reduced critical field strength, denoted by $h_0(\theta)$, is obtained from (2.6) and (2.3) or (2.4) with an equality sign. If

$$x=-h_0(\theta)\sin\alpha>0,$$

then

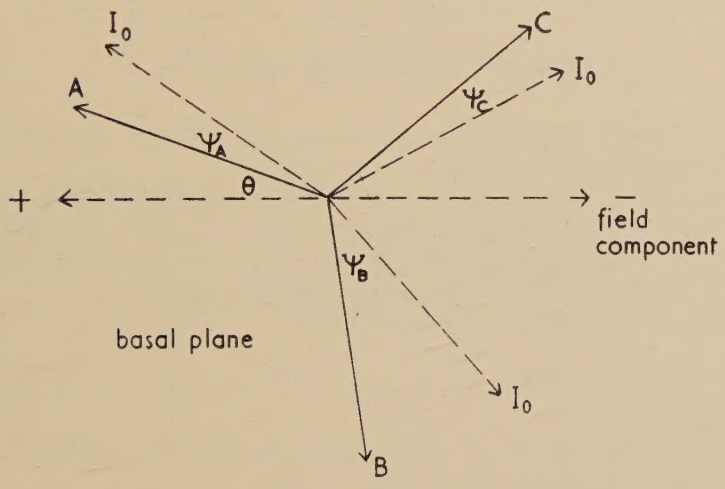
$$\left.\begin{aligned} &\theta=\tan^{-1}(\tfrac{1}{3}t)-\tfrac{1}{3}\tan^{-1}t, \\ \text{where}\quad &t=\tan 3\psi_0=\{(9-4x^2)/(4x^2-1)\}^{1/2}. \end{aligned}\right\} \qquad \ldots\ldots \quad (2.7)$$

With

$$0\leqslant\theta\leqslant\pi/3, \quad 0\leqslant\psi_0\leqslant\pi/6 \quad \text{and} \quad \tfrac{3}{2}\geqslant x\geqslant\tfrac{1}{2}.$$

In the completely demagnetized state the magnetization vectors for one-third the particles with orientation angle θ are along the easy axis A, for one-third along B and for one-third along C, the angle between any two easy axes being $2\pi/3$; see fig. 2 showing the basal plane.

Fig. 2



Angle notation for the basal plane. In the demagnetized state the magnetization vectors I_0 of one-third the particles are parallel to each of the three easy axes A, B and C. On application of a field in the negative direction the magnetization vectors rotate towards the field direction, making angles ψ_A, ψ_B and ψ_C with A, B and C respectively. In the state of remanence after saturation in the positive direction the magnetization vectors are parallel to A. $0\leqslant\theta\leqslant\pi/3$.

When the field component $H\sin\alpha$ is applied, for convenience, in the *negative* direction in the basal plane (see fig. 2) the magnetization vectors rotate towards this direction until, when the field has reached a critical

value, i.e. when $h=HI_0/3K=|h_0(\theta)|$, those originally parallel to A, now making an angle $\psi_A=\psi_0(\theta)$ with A, and those originally parallel to B, now making an angle $\psi_B=\pi/3-\psi_0(\theta)$ with B, rotate discontinuously towards C. After the discontinuous rotations have taken place these two sets of magnetization vectors are parallel to each other and to the magnetization vectors originally parallel to C. On now reducing the field to zero the total magnetization vector rotates continuously back to C. For particles with orientation angles θ and α the component of the remanent magnetization along the direction **H**, $I_R(\theta, \alpha)$, is thus given by

$$\left.\begin{aligned} I_R(\theta,\alpha)/I_0=j_R(\theta,\alpha)&=0, && h<|h_0(\theta)|,\\ j_R(\theta,\alpha)&=\cos(\pi/3-\theta)\sin\alpha, && h\geqslant|h_0(\theta)|,\end{aligned}\right\} \quad . \; . \quad (2.8)$$

$h_0(\theta)$ being given by (2.7), where $x=|h_0(\theta)|\sin\alpha$. If the orientations of the particles are random then the reduced remanent magnetization of the assembly $j_R=I_R/I_0$ is given by

$$\left.\begin{aligned} j_R&=\frac{3}{\pi}\int_0^{\pi/2}\sin^2\alpha\,d\alpha\int_\theta^{\pi/3}\cos\left(\frac{\pi}{3}-\theta\right)d\theta,\\ &=\frac{3}{\pi}\int_0^{\pi/2}\sin\left(\frac{\pi}{3}-\theta\right)\sin^2\alpha\,d\alpha,\\ &=\frac{3}{\pi}h^{-2}\int_0^h\left\{x^2\sin\left(\frac{\pi}{3}-\theta\right)/(h^2-x^2)^{1/2}\right\}dx,\end{aligned}\right\} \quad . \; . \quad (2.9)$$

where $h\equiv|h_0(\theta)|=HI_0/3K$, $x=h\sin\alpha$, and where θ is given as a function of x by (2.7).

The analogous much simpler calculation for particles with two easy directions of magnetization (Stoner and Wohlfarth 1948) has been given by Wohlfarth (1955 b). In the present case the main mathematical difficulty lies in the evaluation of the integral in (2.9).

Since $\frac{1}{2}\leqslant x\leqslant\frac{3}{2}$ it follows from (2.9) that

$$\left.\begin{aligned} j_R&=0, && h<\tfrac{1}{2},\\ j_R&=\frac{3}{\pi}I(h), && \tfrac{1}{2}<h<\tfrac{3}{2},\\ j_R&=\frac{3}{\pi}I\left(\frac{3}{2}\right)+\frac{3\sqrt{3}}{4\pi}\left\{\frac{\pi}{2}-\sin^{-1}(3/2h)+3(4h^2-9)^{1/2}/4h^2\right\}, && h>\tfrac{3}{2},\end{aligned}\right\}$$

where

$$I(\xi)=h^{-2}\int_{\frac{1}{2}}^{\xi}\left\{x^2\sin\left(\frac{\pi}{3}-\theta\right)/(h^2-x^2)^{1/2}\right\}dx, \quad . \; . \; . \quad (2.10)$$

θ being given by (2.7). By using as variable of integration the angle

$$\phi=\frac{\pi}{2}-3\psi_0=\tan^{-1}\{(4x^2-1)/(9-4x^2)\}^{1/2}, \quad . \; . \quad (2.11)$$

it is found that

$$\left.\begin{aligned} j_R &= 0, && h<\tfrac{1}{2},\\ j_R &= \frac{4\sqrt{2}}{\pi}\frac{k}{h^2}J_1(\beta), && \tfrac{1}{2}<h<\tfrac{3}{2},\\ j_R &= \frac{4\sqrt{2}}{\pi}\frac{k}{h^2}J_1\left(\frac{\pi}{2}\right)+\frac{3\sqrt{3}}{4\pi}\left\{\frac{\pi}{2}-\sin^{-1}(3/2h)-3(4h^2-9)^{1/2}/4h^2\right\}, && h>\tfrac{3}{2}, \end{aligned}\right\}$$

where

$$J_1(\eta)=\int_0^{\eta}\{1-k^{-2}\sin^2\phi\}^{1/2}\cos\phi\cos(\tfrac{1}{3}\phi)\,d\phi,$$

and where

$$k=\{(4h^2-1)/8\}^{1/2}=\sin\beta.$$

. . (2.12)

The integral $J_1(\eta)$ cannot be evaluated in closed form except when $h=3/2$; it is, however, possible to express $\cos(\frac{1}{3}\phi)$ over the range required in a more amenable approximate form, viz.

$$\left.\begin{aligned} &\cos(\tfrac{1}{3}\phi)=\mathrm{A}+\mathrm{B}\cos\phi+\mathrm{C}\sin^2\phi,\\ &\mathrm{A}=0{\cdot}8400,\quad \mathrm{B}=0{\cdot}1600,\quad \mathrm{C}=0{\cdot}0267, \end{aligned}\right\} \quad \text{where} \quad . \; . \; (2.13)$$

which gives the required values of $\cos(\frac{1}{3}\phi)$ correctly to at least three significant figures and the value of the integral to an even higher accuracy (for $h=3/2$, the exact value of J_1 is 0·7714, the approximate value using (2.13) is 0·7716). Inserting the approximation (2.13) into (2.12), j_R can be evaluated in terms of elementary and elliptic functions. For $(h-\frac{1}{2})\to 0$ and $h\to\infty$ the following series expressions for j_R are obtained:

$$\left.\begin{aligned} j_R &= \frac{\sqrt{2}k^2}{h^2}\{1-0{\cdot}0133k^2+0(k^4)\}, && k^2=(4h^2-1)/8,\\ &=2\sqrt{2}(h-\tfrac{1}{2})\{1-3{\cdot}007(h-\tfrac{1}{2})+\mathrm{O}(h-\tfrac{1}{2})^2\}, && (h-\tfrac{1}{2})\to 0,\\ j_R &= 3\sqrt{3}/8-0{\cdot}0933h^{-3}-0{\cdot}0225h^{-5}-\mathrm{O}(h^{-7}), && h\to\infty. \end{aligned}\right\} \; . \; (2.14)$$

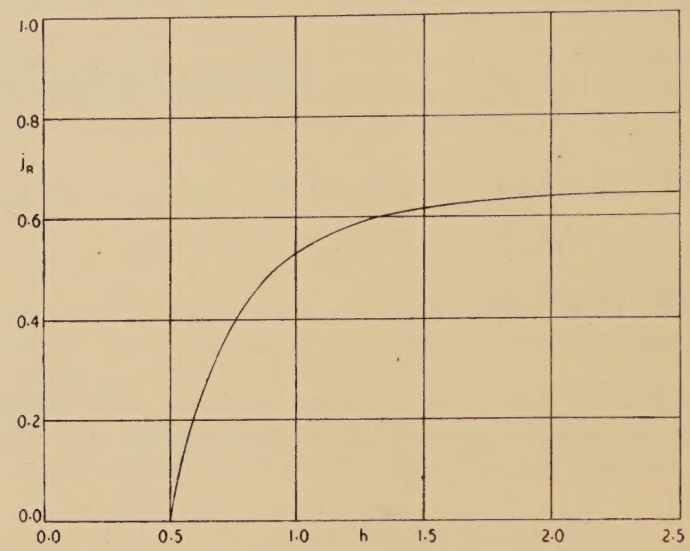
The calculated values of j_R are given in the table and graphically in fig. 3.

The most striking characteristic of the j_R, h curve is the low field foot, extending up to $h=\frac{1}{2}$. Thereafter j_R rises, linearly at first with slope $2\sqrt{2}=2{\cdot}828$, and finally tends asymptotically to the saturation value $3\sqrt{3}/8=0{\cdot}6495$.

The preliminary experiments indicate that the field strength for which j_R rises appreciably has a value in the range 1 to 2×10^3 oersted. From

this the value of K is equal to roughly 10^3 erg cm^{-1}, compared with 5×10^5 erg cm^{-1} for the room temperature value of iron.

Fig. 3



Reduced remanent magnetization.

h	j_R	j_D	h	j_R	j_D
0·50	0·000	+0·650	1·1	0·557	−0·424
0·52	0·053	0·564	1·2	0·579	0·477
0·54	0·101	0·484	1·3	0·596	0·518
0·56	0·143	0·411	1·4	0·608	0·550
0·58	0·182	0·342	1·5	0·618	0·572
0·6	0·216	0·280	2·0	0·637	0·620
0·7	0·346	+0·032	2·5	0·643	0·635
0·8	0·430	−0·140	3·0	0·646	0·641
0·9	0·487	0·264	∞	0·650	−0·650
1·0	0·527	−0·355			

Dependence on reduced field strength $h=HI_0/3K$ of the reduced remanent magnetization $j_R=I_R/I_0$ and the reduced remanence after demagnetization $j_D=I_D/I_0$. For $h<\frac{1}{2}$, $j_R=0$, $j_D=0{\cdot}650$.

§ 3. Remanence after Demagnetization

In the second type of experiment the powder is reduced to the remanent state from complete saturation. A reverse field H is then applied and the remanent magnetization I_D measured after removal of the field.

In fig. 2 showing the basal plane the positive direction is that of initial saturation. On removing the field required for saturation the magnetization vectors of particles with orientation angle θ turn towards A, the nearest easy axis of magnetization. After now applying and then removing a field with component $H \sin\alpha$ in the negative direction the magnetization vectors will be parallel to the easy axis C, that nearest to the reverse field direction, if H is larger than the critical value, depending on θ and α and given by (2.7). Hence the reduced remanent magnetization $j_D = I_D/I_0$ is given by

$$\left.\begin{aligned} j_D &= \frac{3}{\pi}\int_0^{\pi/2}\left\{\sin\theta - \sin\left(\frac{\pi}{3}-\theta\right)\right\}\sin^2\alpha\, d\alpha \\ &= j_D' - j_R. \end{aligned}\right\} \quad \ldots \quad (3.1)$$

Here j_R has been evaluated in § 2 and

$$\left.\begin{aligned} j_D' &= \frac{3}{\pi}h^{-2}\int_0^h \{x^2 \sin\theta/(h^2-x^2)^{1/2}\}\, dx, \\ \text{where} \qquad h &\equiv |h_0(\theta)| = HI_0/3K, \qquad x = h\sin\alpha, \end{aligned}\right\} \quad \ldots \quad (3.2)$$

and where θ is given as a function of x by (2.7). By using the same transformation as that leading to (2.12) it is found that

$$\left.\begin{aligned} j_D' &= 3\sqrt{3}/8, && h < \tfrac{1}{2}, \\ j_D' &= \frac{3\sqrt{3}}{4\pi}\{\sin^{-1}(1/2h) + (4h^2-1)^{1/2}/4h^2\} - \frac{4\sqrt{2}}{\pi}\frac{k}{h^2}J_2(\beta), && \tfrac{1}{2} < h < \tfrac{3}{2}, \\ j_D' &= \frac{3\sqrt{3}}{4\pi}\{\sin^{-1}(1/2h) + (4h^2-1)^{1/2}/4h^2\} - \frac{4\sqrt{2}}{\pi}\frac{k}{h^2}J_2\left(\frac{\pi}{2}\right), && h > \tfrac{3}{2}, \end{aligned}\right\}$$

where

$$J_2(\eta) = \int_0^\eta \{1-k^{-2}\sin^2\phi\}^{1/2}\cos\phi\cos\left(\frac{\pi+\phi}{3}\right)d\phi,$$

and where

$$k = \{(4h^2-1)/8\}^{1/2} = \sin\beta.$$

. . (3.3)

By using the approximate expression

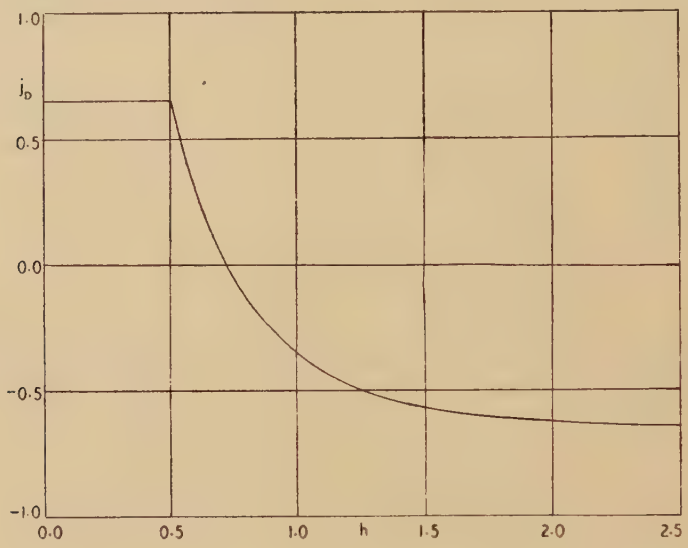
$$\left.\begin{aligned} \cos\left(\frac{\pi+\phi}{3}\right) &= \mathrm{A} + \mathrm{B}\cos\phi + \mathrm{C}\cos^2\phi + \mathrm{D}\sin\phi + \mathrm{E}\sin\phi\cos\phi, \\ \text{where} \quad \mathrm{A} &= 0{\cdot}2874, \quad \mathrm{B} = 0{\cdot}3216, \quad \mathrm{C} = -0{\cdot}1090, \quad \mathrm{D} = -0{\cdot}2874, \\ & \qquad\qquad\qquad\qquad\qquad\qquad \mathrm{E} = 0{\cdot}0036, \end{aligned}\right\} \quad (3.4)$$

which gives sufficient accuracy over the range required, the integral J_2 may be evaluated and hence j_D' and, from (3.1), j_D calculated. The values of j_D are given in the table and graphically in fig. 4.

It is seen that the reduced remanence after demagnetization retains its saturation value $3\sqrt{3}/8 = 0{\cdot}6495$ until, when h exceeds $\frac{1}{2}$, j_D decreases,

linearly at first with negative slope $3\sqrt{2}=4{\cdot}243$, passes through zero when $h=0{\cdot}720$ and finally tends asymptotically to its saturation value $-0{\cdot}6495$ in the reverse direction. The preliminary experiments indicate that the field at which $j_D=0$ is rather larger than that calculated and shown in fig. 4.

Fig. 4



Reduced remanence after demagnetization.

Acknowledgments

It gives me great pleasure to acknowledge the help obtained by discussions with Professor P. M. S. Blackett, F.R.S., and Mr. G. Haigh about their experiments which stimulated the present work, and with Dr. A. N. Gordon about the calculations.

References

Anderson, P. W., Merritt, F. R., Remeika, J. P., and Yager, W. A., 1954, *Phys. Rev.*, **93,** 717.

Morin, F. J., 1950, *Phys. Rev.*, **78,** 819.

Néel, L., 1949, *Ann. Phys., Paris*, **4,** 249 ; 1953, *Rev. Mod. Phys.*, **25,** 58 ; 1955, *Advances in Physics*, **4,** 191.

Néel, L., and Pauthenet, R., 1952, *C.R. Acad. Sci., Paris*, **234,** 2172.

Nicholls, G. D., 1955, *Advances in Physics*, **4,** 113.

Shull, C. G., Strauser, W. A., and Wollan, E. O., 1951, *Phys. Rev.*, **83,** 333.

Stoner, E. C., and Wohlfarth, E. P., 1948, *Phil. Trans. Roy. Soc.* A, **240,** 599.

Wohlfarth, E. P., 1955 a, *Proc. Roy. Soc.* A (in the press) ; 1955 b, *Research*, **8,** S42.

CXXIX. *Radioactive ^{186}Tantalum*

By A. J. Poë*
Atomic Energy Research Establishment, Harwell†

[Received August 10, 1955]

Abstract

The new radioactive isotope ^{186}Ta has been prepared and its decay characteristics determined as follows :—

Half-life :	$10{\cdot}5 \pm 0{\cdot}5$ min
Maximum β-energy :	$2{\cdot}2_2$ MeV
Conversion-electron energies :	$\lesssim 0{\cdot}15$ MeV
γ-ray energies :	125, 200, 300, 410, 510, 610, 730, 940 and possibly $\sim$1150 kev

The mass assignment suggested by its production by (n, p), but not by (γ, p), reactions on tungsten was confirmed by experiments using tungstic acid enriched in ^{186}W.

§ 1. Introduction

Of the neutron excess tantalum isotopes which can be made by fast neutron irradiation of tungsten, only ^{186}Ta has not previously been reported. This paper describes the production of this isotope and the examination of its decay characteristics.

§ 2. Experimental Techniques

The neutron excess tantalum isotopes were produced by irradiation of tungstic acid with fast neutrons generated by proton bombardment of a beryllium target in the Harwell cyclotron. The tantalum isotopes were quickly separated by coprecipitation with ferric hydroxide from the solution of tungstic acid in sodium hydroxide. Suitable samples for counting were obtainable within about seven minutes of the end of the irradiation. 10-minute irradiations of 5–20 g of tungstic acid with neutrons of energies up to 50 MeV produced sources, with intensities of about 1 μc, for the examination of the decay characteristics as described below.

§ 3. Results

(i) *Production and Mass Assignment*

After irradiation of tungstic acid with neutrons of energies up to 20 MeV, so that tantalum isotopes were produced mainly by (n, p) reactions on the tungsten, a half-life of about 10 minutes was observed in the

* Now at the Chemistry Dept., Imperial College, London, S.W.7.
† Communicated by the Author.

beta-decay of the tantalum sources in addition to a 9-hour half-life and longer-lived activities. This 10-minute half-life was also observed when low energy conversion electrons from any 16-minute ^{182m}Ta present were absorbed by 60 mg/cm^2 of aluminium. The decay of the more intense sources produced by irradiations with neutrons of energies up to 50 Mev also showed the 10-minute half-life when the particles counted had passed through 500 mg/cm^2 of aluminium, which was more than enough to absorb any 0·6 Mev beta-particles from ^{182m}Ta (Wilkinson 1950). Spectrographic analysis of the tungstic acid, and test irradiations of iron, showed that the 10-minute activity could not have been due to contamination by 9-minute ^{53}Fe.

The 10-minute activity was not observed after irradiation of tungstic acid with an intense beam of x-rays, of energies up to 28 Mev, generated by a linear electron accelerator. Its production by (n, p), but not by (γ, p), reactions on tungsten suggests that the 10-minute activity should be assigned to ^{184}Ta or ^{186}Ta but ^{184}Ta is known to have an 8·7-hour half-life (Butement and Poë 1955). That the 10-minute activity must be assigned to ^{186}Ta was confirmed by experiments with tungstic acid enriched in ^{186}W. In one experiment equal masses of normal and enriched tungstic acid were irradiated by neutrons with energies of up to 20 Mev. The irradiations were monitored with the 12·8-hour activity induced in copper foils placed between the neutron source and the tungstic acid target. Tantalum activities were separated from the two samples under closely similar conditions and the decays of the beta-particles passing through 30 mg/cm^2 of aluminium were followed. The ratio of the yield of the 10-minute activity from the enriched tungstic acid to that from the normal tungstic acid was 6·4 : 1. Since the enrichment factors of the tungsten isotopes of mass 180, 182, 183, 184 and 186 were 0·09, 0·02, 0·03, 0·05 and 3·4 respectively, relative to natural tungsten, the increase in yield of the 10-minute activity was closest to the enrichment of ^{186}W.

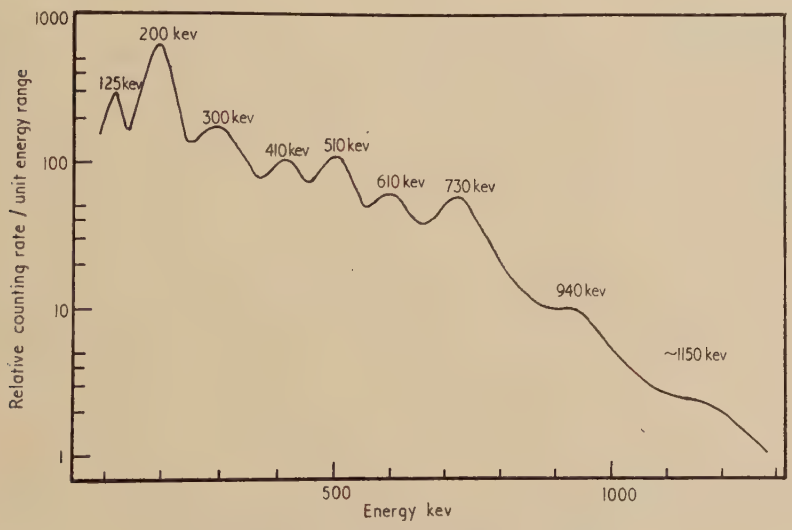
In the other experiment the ratio of the intensity of the 10-minute activity to that of the 50-minute ^{185}Ta activity produced by (n, pn) reactions on ^{186}W was measured for tantalum samples prepared by irradiation of normal and ^{186}W-enriched tungstic acid with neutrons of energies up to 50 Mev. Low energy electrons from the tantalum samples were absorbed by 60 mg/cm^2 of aluminium. In both cases the ratio was the same indicating that the 10-minute tantalum activity was also produced by fast neutron bombardment of ^{186}W.

The 10-minute activity can therefore be unambiguously assigned to ^{186}Ta.

(ii) *Radiation Characteristics*

The half-life was obtained by measurement of the decay of beta-particles passing through 500 mg/cm^2 of aluminium which absorbed all particles from isotopes other than the 10-minute ^{186}Ta and the 50-minute ^{185}Ta. The half-life found as the average of eight determinations was 10·5 minutes, all the results lying within 0·5 minutes of this value.

The aluminium absorption curve of the beta-particles from short-lived tantalum activities indicated a maximum beta-particle energy of 2·3 MeV together with a low energy component of about 0·15 MeV with approximately the same intensity. Beta–gamma coincidence absorption experiments, using a Geiger counter and NaI crystal scintillation counter, showed that the high energy beta-disintegration is in coincidence with gamma-rays. While some of the low energy particles must have been conversion electrons from ^{182m}Ta many were conversion electrons in coincidence with the 10-minute ^{186}Ta beta-disintegration. This was shown by the aluminium absorption of those particles entering one end-window Geiger counter in coincidence with beta-particles entering a similar counter from which conversion electrons and low energy beta-particles were excluded by aluminium absorbers. Spurious coincidences caused by scattering of a particle from one counter into the other were almost completely eliminated by mounting the source over a small hole in a brass plate placed between the two counters. More than 90% of the coincidences were due to particles with energies up to about 0·15 MeV. Therefore most of the conversion electrons from the various conversion processes possible, in view of the gamma-ray spectrum described below, have energies of about 0·15 MeV or less.



Gamma-ray energy spectrum of ^{186}Ta

The beta-particle energy spectrum was examined from 1·25 MeV upwards with an anthracene crystal scintillation spectrometer together with a 30-channel pulse analyser. The spectrometer was calibrated with conversion electrons following the decay of ^{113}Sn and ^{137}Cs and beta-particles from ^{65}Ni (2·10 MeV) and ^{90}Y (2·24 MeV). Fermi beta-disintegration functions given by Feister (1950) were used for the Fermi plots.

The maximum beta-particle energy of ^{186}Ta was found to be $2{\cdot}2_2$ Mev as the average of twelve closely agreeing results.

The gamma-ray spectrum of ^{186}Ta was examined with a NaI crystal scintillation spectrometer and a 30-channel pulse analyser. The spectrum was covered in several overlapping stages and at two different amplifier gain settings, two or three successive counts of the short-lived activity being taken from each sample and corrections for longer-lived activities being applied. The spectrum up to 300 kev was obtained from samples prepared by irradiation of ^{186}W-enriched tungstic acid to avoid interference from the 180 kev gamma-ray from ^{182m}Ta. The results were all combined to give the spectrum shown in the figure. No 60 kev K x-ray component was observed because of the very large background intensity of this energy from 50-minute ^{185}Ta. The spectrum is complex and may be interpreted to indicate gamma-rays, or groups of gamma-rays, of energies close to 125, 200, 300, 410, 510, 610, 730 and 940 kev. After allowing for Compton absorption and for photo-electric absorption efficiency the relative intensities of the above gamma-rays were found to be very approximately 2·5 : 10 : 2·5 : 2 : 4·5 : 4·5 : 6·5 : 1·5. The 125 kev component is probably at least partly due to the transition of that energy which has also been observed following electron capture by ^{186}Re (Metzger and Hill 1951, Steffen 1951) and Coulomb excitation of ^{186}W (McClelland, Mark and Goodman 1954).

The higher energy component at about 1150 kev was not clearly resolved and not very reproducible. It was probably due in part to simultaneous absorption in the crystal of two lower energy gamma-rays. Experiments with the source 1 in. from the crystal suggest, however, that there is a genuine gamma-ray, or group of gamma-rays, at 1100–1200 kev with intensity $\lesssim$1 relative to the 730 kev gamma-ray intensity of 6·5.

Acknowledgments

Acknowledgments are gratefully made to Mr. R. B. Thomas for much practical assistance and to the members of the cyclotron group for performing the irradiations.

References

Butement, F. D. S., and Poë, A. J., 1955, *Phil. Mag.*, **46**, 482.
Feister, I., 1950, *Phys. Rev.*, **78**, 375.
McClelland, C. L., Mark, H., and Goodman, C., 1954, *Phys. Rev.*, **93**, 904.
Metzger, F. R., and Hill, R. D., 1951, *Phys. Rev.*, **82**, 646.
Steffen, R. M., 1951, *Phys. Rev.*, **82**, 827.
Wilkinson, G., 1950, *Phys. Rev.*, **80**, 495.

CXXX. *On the Linear Work Hardening Rate of Face-centred Cubic Single Crystals*

By J. Friedel
Centre de Recherches Métallurgiques de l'Ecole des Mines de Paris *

[Received August 16, 1955]

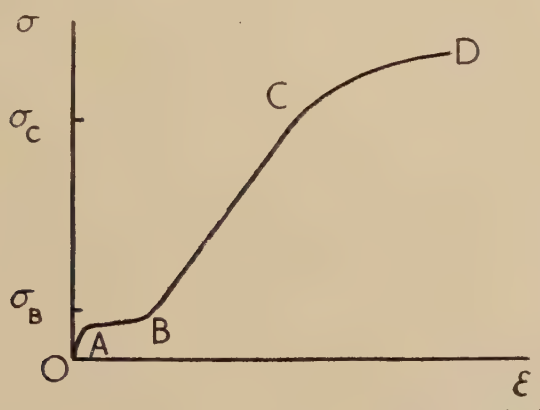
Summary

The linear strain hardening observed after the easy glide region in face-centred cubic single crystals at sufficiently low temperature is explained by assuming that all the Frank–Read sources of the active slip system are in action, and form small active slip planes bounded on all four sides by Cottrell barriers. The initial easy glide region should occur before all the barriers are formed ; the parabolic strain hardening appearing in the final stages would be related to a thermally activated breaking of the piled-up dislocations through the Cottrell barriers.

§ 1. Introduction

The stress–strain curves observed by traction of face-centred cubic single crystals exhibit, after the initial easy glide stage AB (fig. 1), a linear part BC which has as yet received no satisfactory explanation.

Fig. 1



Typical stress–strain curve of a face-centred cubic single crystal

It is usually agreed that easy glide occurs as long as the dislocations of the principal slip system can develop without much hindrance, and, at least in face-centred cubic crystals, either go out of the crystal or

* Communicated by the Author.

compensate each other's stress field in deformation bands. Also greater hardening rates occur when some ' secondary ' slip on another system hinders the principal slip system too much. In agreement with this idea, easy glide is in evidence particularly under conditions such that a principal slip system is much more favoured than the others, the extreme case being that of hexagonal metals (Mott 1953, Cottrell 1953). The second stage BC actually occurs before the axis of traction has reached the [100]–[111] symmetry zone, where two slip systems are equally favoured. But this is usually explained by inhomogeneities in the stresses due to the dislocations developed in the principal slip system ; or, more macroscopically, by a lower work hardening in the secondary slip systems (Röhm and Kochendörfer 1948, Stroh 1953).

It is now clear that the more rapid work hardening rate, which sets in after the easy glide stage, begins with an almost perfectly *linear* stage BC (fig. 1) in all the face-centred cubic crystals investigated and whatever the initial orientation of their axis of traction, if the temperature of deformation is sufficiently low. The metals and alloys studied so far are aluminium, nickel, copper, silver, gold, alpha brass and Cu_3Au (c.f. Masima and Sachs 1928, von Göler and Sachs 1929, Sachs and Weerts 1930, 1931, Blewitt and Köhler 1950, Jaoul 1955, Kuhlmann-Wilsdorf and Wilsdorf 1953, Blewitt 1953, 1955, Rosi 1954, Staubwasser 1954, Kocks 1954, Cottrell and Stokes 1955, Diehl, Mader and Seeger 1955). *The rate of work hardening* $(d\sigma/d\epsilon)$ *is about* $\frac{1}{2}\ 10^{-2}\mu$, increasing slightly for axes of traction with orientations varying from [110] to the [100]–[111] zone ; but the factor of proportionality to the shear modulus μ *does not seem to vary appreciably with temperature, nor with the nature of the metal.* On the other hand, the last stage CD (fig. 1), with negative curvature, appears at stresses lower for higher temperatures and, at the same temperature, lower for aluminium than for copper. These two effects combine so as to reduce the linear stage BC to practically nothing in aluminium at room temperature, and to produce twinning without any third stage CD in copper at 4°K.

Some current theories of work hardening (Taylor 1934, Mott 1953) predict parabolic stress–strain curves, and cannot therefore explain the linear stage BC. This note suggests a possible mechanism for this behaviour, and for the variation of the parabolic stage CD with temperature and with the nature of the metal. The main difference with the previous theories will be to assume, in the linear stage, that piled up groups of dislocations are fixed in number and position, and increase only in strength with deformation. Previous theories assumed on the contrary that the dislocations were dispersed (Taylor) or piled up in groups of fixed strength (Mott) but increasing in number with deformation.

§ 2. Cottrell Barriers

Work hardening is certainly due to an accumulation of dislocations in the crystal. This can impede further movements of dislocations in two ways (c.f. Mott 1953, Cottrell 1953, Seeger 1954) :

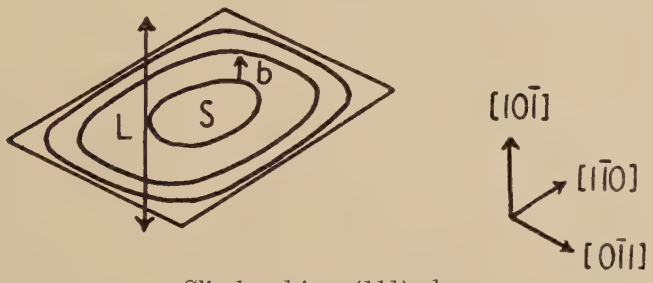
1. Multiplication of the forest of screw dislocations through which the moving dislocations have to cut.

2. Increase in the elastic distortion of the crystal, which hinders the movement of dislocations.

The creation of jogs necessary in the first process is a thermally activated process, with quite a small activation energy for metals like aluminium ; it should lead to a work hardening strongly dependent on temperature, contrary to what is observed in the linear part BC. A dependence on temperature proportional to that of the elastic constants, as is predicted by the *second* mechanism if no recovery occurs, seems in better agreement with observation.

For this second mechanism to be satisfactory, one must find a reason, other than cutting through the forest of screws, for dislocations to remain in the crystal. In single crystals, this can be achieved only by elastic interaction with dislocations of the same slip system and opposite sign, as in deformation bands, or by piling up dislocations in front of Cottrell barriers formed with dislocations of other slip systems. Deformation bands cannot harden the crystal very much, for the long range actions of their dislocations with opposite signs compensate each other (Mott 1953). They form indeed in the easy glide stage. *The beginning of the linear stage* BC *must therefore be connected with the appearance of Cottrell barriers.*

Fig. 2



Slip band in a (111) plane.

A *great number* of such barriers must be formed from the beginning of the linear stage. For, as will now be shown, *every Frank–Read source belonging to the principal slip system must be active* ; and *every active source must form at least four barriers in its glide plane.*

The first conclusion is reached by noting that if, in the linear stage, dislocations cut through the Frank–Read net without difficulty even at low temperatures, they must be acted upon by stresses large enough to make active most of the Frank–Read sources of the principal slip system. This is not surprising, as the stresses applied in the linear stage are always much larger than $10^{-4}\,\mu$.

An active source S can, on the other hand, form barriers along the two [110] directions which are not parallel to its own Burgers vector (c.f. Appendix A). It can therefore surround itself with four barriers, forming the parallelogram represented in fig. 2.

Now, if hardening is only due to piled-up groups in front of Cottrell barriers, each active source *must* be surrounded by such a parallelogram. Otherwise, loops from S could escape to the surface of the crystal, and produce an infinite amount of deformation under a constant stress.

§ 3. Short Slip

The *size* L of these parallelograms may be deduced from the two following considerations :

1. As shown in Appendix A, the density $N_{S'}$, of sources S′ able to form barriers of a given direction with the sources S is roughly the same as the density N_S of S sources. To build all the parallelograms required, *all the possible* S′ *sources must therefore be used*, and *each of them must build* $2N_S/N_{S'} \simeq 2$ *barriers with active sources* S. That loops from a given source can build barriers with loops from several other sources is shown clearly by Jacquet's observations on alpha brass (1954).

2. Cottrell barriers are formed readily, as soon as two intersecting slip systems of suitable slip directions are active. This is again shown by Jacquet's pictures. This must be due to the attraction between dislocations able to form barriers readily (c.f. Appendix A) and to the considerable amount of energy gained by forming such a barrier (usually several electron volts per interatomic distance). We can therefore assume that, from the start of the linear stage, *each active source* S *has built up barriers with the nearest* S′ *sources available*. In these conditions, the size L of the parallelograms is the average distance between S′ sources :

$$L \simeq (N_{S'})^{-1/3}.$$

The linear stage should therefore set in when the orientation of the axis of traction and the hardening due to easy glide are such as to activate the sources S′ of the required slip system to form barriers. It should be a region of *short slip bands*, of *constant number* and *length*, distributed *at random* through the crystal.

For a *well annealed* crystal, we may assume that the possible Frank–Read sources build up a three-dimensional net of size l, where the various slip systems are equally represented (Mott 1953). Then $N_{S'} \simeq N_S$ and, when one of the 12 slip systems is active, $N_S \simeq 1/12l^3$. With $l = 10^{-4}$ to 10^{-3} cm, *the lengths and distances of the slip bands should be of about* 10^{-3} cm.

These conclusions, deduced from an analysis of the stress–strain curves, are in full agreement with observations made independently on copper by Blewitt (1955). Blewitt examined the slip lines produced by an elongation of a few per cent, after prior traction into the linear stage followed by electropolishing. The surface was thus free in these experiments from the slip lines produced in the easy glide stage. It exhibited all the characteristics predicted, except apparently for some reduction in size of the slip lines with increasing strain. As pointed out by Blewitt himself, however, the dispersion of the statistical measurements made this last

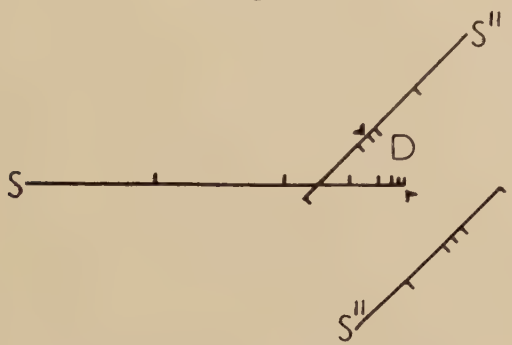
effect uncertain. As expected too, the slip lines in the easy glide stage were quite different, strong lines, much longer and fewer in number, appearing in what is, according to Diehl, Mader and Seeger (1955), a background of fine slip.

The hardening of polygonized crystals will be considered in § 5.

§ 4. Linear Hardening

In our model, hardening may be attributed to the internal stresses acting on the sources S. They have three possible origins: the back stress σ_1 of a piled-up group on its emitting source; a 'relaxation' σ_1' of this stress, described below; and the stress σ_2 on the source due to all the other piled-up groups. If, as assumed here, all the piled-up groups have similar numbers n of loops, n increasing uniformly with strain, these three stresses are proportional to n. This explains a *linear* hardening, for the strain varies linearly with n.

Fig. 3



Relaxation around a piled up group.

Treating the n loops of a piled up group (fig. 2) as concentric circles, we have (Leibfried 1951)

$$\sigma_1 \simeq \frac{n\mu b}{2L} \qquad (1)$$

As the strain due to the piled up groups is

$$\epsilon \simeq n N_S L^2 b, \qquad (2)$$

the first term, taken alone, would give a hardening rate

$$\frac{d\sigma_1}{d\epsilon} \simeq \frac{\mu}{2N_S L^3}. \qquad (3)$$

For a *well annealed* crystal, where $N_S \simeq L^{-3}$ is of the order of N_S, $\frac{d\sigma_1}{d\epsilon} \simeq \frac{1}{2}\mu$, *which is too large by a factor about* 10^2.

We attribute the main part of this discrepancy to a relaxation of σ_1, produced by the neighbouring Frank–Read sources. As pointed out

by Mott (1953), as soon as a piled up group has more than a few dislocations, its stress field is strong enough to set into action all the neighbouring sources S″ (fig. 3). These emit loops of dislocations in such a way as to reduce the stresses around the piled up groups.

Practically *all* the Frank–Read sources can be used for this relaxation. In a well annealed crystal and if each source S″ contributes loops to the two nearest piled up groups, there will be 12–1=11 sources S″ per piled up group, if one of the 12 slip systems is active. The average distance D over which relaxation of a piled up group occurs should therefore be of the order of a fraction $\frac{1}{2\times 11}$ of the distance between S″ sources, so that $D\simeq(l/20)$, if l is the size of the Frank–Read net. We may reasonably assume that the relaxation transforms the effect of the piled up group on its S source into that of a *dipole* of average strength nD. The back stress on S is thus reduced from σ_1 to $\sigma_1+\sigma_1'=(D/L)\sigma_1$; and, as $L\simeq l\times(12)^{1.3}$,

$$\frac{D}{L}\simeq\frac{1}{20\times(12)^{1/3}}\simeq\frac{1}{50}.$$

The contributions to the strain ϵ from the various S″ sources during relaxation probably cancel each other. Thus ϵ is still given by (2). And *the relaxation reduces the hardening rate by the factor* $D\,L\simeq 2\times 10^{-2}$, of the right order of magnitude to explain the rates observed :

$$\frac{d\sigma}{d\epsilon}=\frac{d(\sigma_1+\sigma_1')}{d\epsilon}\simeq\frac{\mu D}{2L}\simeq 10^{-2}\,\mu. \qquad (4)$$

The stress σ_2 due to the other (relaxed) piled up groups has been neglected in the hardening rate, for the following reason. These groups occur in pairs of opposite sign, and therefore cancel their effects, except for the group nearest to the S source considered, at a distance, say, λ (fig. 6). Thus, for a well-annealed crystal,

$$\sigma_2\simeq\pm\frac{nubD}{4\pi\lambda^2}. \qquad (5)$$

λ is of the order of the distance between S sources, thus L, according to § 3 ; and σ_2 is of the order of $\sigma_1+\sigma_1'$, but with an equal probability of being positive or negative. If, therefore, σ is increased by $d\sigma$, the sources for which the corresponding increase $d\sigma_2$ is positive or negative will be present in equal number ; they will emit respectively dn_+ or dn_- loops, so that

$$d\sigma=d\sigma_2+\frac{\mu bD}{2L^2}dn_+=-d\sigma_2+\frac{\mu bD}{2L^2}dn_-$$

$$=\frac{\mu bD}{2L^2}\,\frac{dn_++dn_-}{2}=\frac{\mu D}{2L}d\epsilon.$$

The stresses σ_2 due to the other piled-up groups should not therefore effect the hardening rate, which is the same as (4). They may, on the

other hand, play an important role in stabilizing the piled up groups : one has to explain why the piled up groups remain in the crystal when the applied stress is removed. Mott (1953) proposed that piled up groups could be locked by Cottrell barriers formed on their rear with loops emitted by S'' sources during relaxation. This process is however not possible with the model used here, for all the S' type of sources able to form barriers with the dislocations of the piled-up groups have been used up to form the parallelograms on which piling up occurs (§ 2). The loops emitted by the S'' sources during relaxation interact therefore only elastically with the piled up groups*. They reduce their energy, thus making them more easily stabilized by any frictional force such as the interactions σ_2 with other piled up groups. This possible stabilization is, however, not yet clear in detail.

§ 5. Hardening of Polygonized Crystals

The equations of the last section can be put into a more general form, which shows that the hardening rate should *not* depend very much on the distribution and size of the Frank–Read sources in the metal.

Let N_S, $N_{S'}$ and $N_{S''}$ be the densities of S, S' and S'' sources in the crystal (S active sources ; S' sources able to form barriers readily with S sources ; S'' sources other than S). D is, according to § 3, a fraction $N_S/2N_{S''}$ of the average distance $N_{S''}^{-1/3}$ between S'' sources. Thus

$$D \simeq N_S/2(N_{S''})^{4/3}$$

and

$$\frac{d\sigma}{d\epsilon} \simeq \frac{\mu}{4L^4(N_{S''})^{4/3}},$$

with $L \simeq (N_{S'})^{-1/3}$ if $N_{S'} \leqslant N_S$ and $N_S/N_{S'}^{4/3}$ if $N_{S'} \geqslant N_S$. The value of L for $N_{S'} \geqslant N_S$ comes from the fact that there will be $N_{S'}/N_S$ sources S' able to form a barrier with each source S, and their average distance is $N_{S'}^{-1/3}$. The value of L for $N_{S'} \leqslant N_S$ was obtained in § 3. Finally, if the sources $S_{1'}$ and $S_{2'}$ able to form barriers along the two directions [110] and [011] of fig. 2 have unequal densities, thus give unequal slip distances $L_1 < L_2$, the factor L^4 in the hardening rate is to be replaced by $L_1^3 L_2$. For the variation of σ only depends on the action of the piled up groups at the shorter distance L_1 ; and L_2 enters into the equation only through the area $L_1 L_2$ of the loops.

These equations were obtained assuming the S, S' and S'' sources distributed fairly *uniformly* in the crystal. With this assumption, some interesting results are obtained.

1. The hardening rate $d\sigma/d\epsilon$ depends only on the *densities* of S, S' and S'' sources, not on their exact geometrical arrangement. Thus it is the same whether the crystal is polygonized or well annealed, provided N_S, $N_{S'}$, $N_{S''}$ are the same.

---

* They may form Cottrell barriers between themselves.

2. The hardening rate depends only on the *ratios* of these densities, not on their absolute magnitude : it does not depend on the size of the Frank–Read net.

3. In the general case where $N_{S'} \leqslant N_S$, the hardening rate *does not depend on the density of the active* S *sources*. Thus a (finely) polygonized structure where the walls of the blocks are built up with dislocations of the *active* slip system should have the same hardening rate as a well annealed crystal with the same densities of dislocations in the other slip systems, although its density N_S of S sources is much larger. This is because the lowering of the hardening rate $d\sigma_1/d\epsilon$ due to the piled up groups (cf. (3)) is compensated by a decrease in relaxation $(\sigma_1+\sigma_1')/\sigma_1=D/L$, due to fewer S'' sources per piled up group. There are, however, many more slip lines than in the well annealed crystal, for the number of slip lines is given by N_S.

This is possibly Kuhlmann-Wilsdorf and Wilsdorf's ' fine slip ', observed on mono- and polycrystals of high purity aluminium after fairly large deformations (cf. Kuhlmann-Wilsdorf, van der Merwe and Wilsdorf 1952). It consists of parallel slip lines, of length L from 5 to 20×10^{-4} cm and with distance X from 3 to 5×10^{-6} cm apart. It could be originated by a polygonized structure where the blocks of the mosaic structure, of size $l\simeq L/\sqrt[3]{12}$, are covered by an array of $l/X\simeq100$ parallel dislocations belonging to the active slip system. High purity metals are known to present readily a polygonized structure of that type, when heat treated for recrystallization (Talbot, de Beaulieu and Chaudron 1953, Friedel, Boulanger and Crussard 1955). Finally the number n of loops deduced from the height of the slip lines in the fine slip is near to $L\sigma/\mu b$, in approximate agreement with eqn. (1). There is therefore, as expected, little relaxation of the stress σ_1 of the piled up groups in that case :

$$\sigma/\sigma_1=(\sigma_1+\sigma_1')/\sigma_1=D/L\simeq 1.$$

4. The only case where $N_{S'} \gg N_S$ is in a polygonized structure where the dislocations of the walls belong to one of the two types of S' sources. The equations given above predict then a very short slip distance L_1 in that direction, and as trong hardening rate, varying as $(N_{S'})^{16/3}$. It would be interesting to check this point. However when the active slip system is sufficiently hardened in that way, other systems become probably very active and distort the picture.

5. All other cases of polygonization should lead to hardening rates of the same order as for well annealed crystals, if somewhat smaller. This is because the density of S'' sources is large anyway : to lead to a noticeable decrease in the hardening rate, the polygonization must increase the density of sources, in one of the 11 slip systems making up the S'' sources, by a factor much larger than 10. This is not a very likely occurrence.

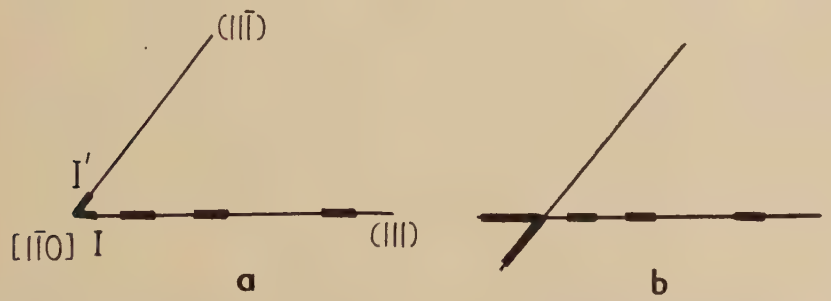
§ 6. Breaking Up of Cottrell Barriers

The ' parabolic ' hardening CD (fig. 1) occurs at a stress σ_C which varies strongly with temperature, thus indicating some thermally activated process. We suggest that it occurs *when the stress concentration in front of the piled up groups is large enough to force, with the help of temperature, some of its dislocations through a Cottrell barrier*. This would extend the slipped region around a source S further than the parallelogram of fig. 2, and give rise to a hardening rate lower than the linear stage.

As an experimental confirmation of this point of view, we may refer to Diehl, Mader and Seeger (1955), who observed on several metals (Al, Cu, Ni) extended and multiple slip bands appearing in the parabolic stage, in contrast with the fine slip of the linear stage. More systematic observations of the Blewitt type (1955) would certainly be welcome.

There are of course several possible geometrical configurations for the piled up groups and the barriers, so various possible mechanisms for breaking through a barrier. Figure 4 represents two possible configurations.

Fig. 4



Two possible geometrical configurations for a group of dislocations piled up in front of a Cottrell barrier.

A complete discussion will be given by Stroh (1955). We shall only describe here a mechanism connected with the configuration of fig. 4(a), which we believe to be the most favourable for breaking through.

Let us assume that, by thermal agitation, dislocation I has recombined with the barrier over a certain length h. Then the whole barrier has recombined over this length ; for the upper half dislocation I' has a Burgers vector $(b/3\sqrt{2})$ [112] which is now normal to the Burgers vector $(b\sqrt{2}/3)$ [111] of the rest of the barrier, and is therefore no longer repelled by it (fig. 5).

Now, if h is larger than the critical length of a Frank–Read source under an applied stress $n\sigma$, that is if *

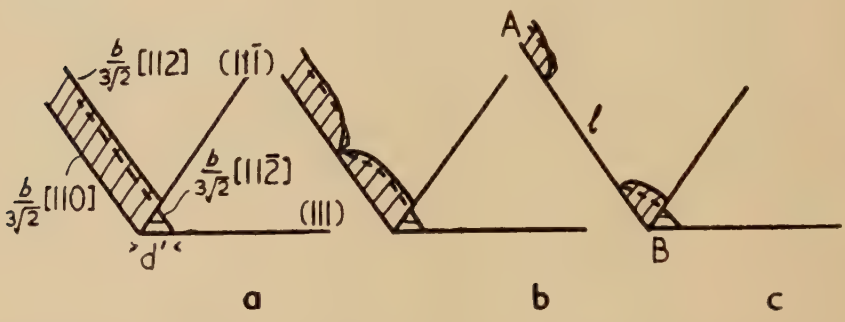
$$n\sigma \geqslant \frac{\mu b}{2\pi K h} \ln \frac{h}{b_0}, \qquad \ldots\ldots (6)$$

* h may actually be somewhat smaller, as will be explained by Stroh (1955).

with $b_0 \simeq b$ and $K \simeq 0{\cdot}8$ for the orientation of the barrier (cf. Cottrell 1953), the recombined portion of the barrier, with a Burgers vector $(b/\sqrt{2})$ [110], can bend in its (001) glide plane into a loop able to expand. There is therefore a stable final state where the dislocations have broken through the barrier, and which can be reached by thermal activation.†

The activation energy is given by the difference of energy between the states of figs. 5(*a*) and (*c*). It is the sum of the energy U_1 to recombine the barrier over a length h, and that U_2 to produce a 'constriction' such as in fig. 5(*b*). According to Stroh (1954, 1955),

Fig. 5



Example of Cottrell barrier : (*a*) normal state ; (*b*) constricted ; (*c*) recombined.

$$U_1 = \frac{\mu b^2 h}{18\pi K}\left(\ln\frac{d'}{b_0} - 1\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and

$$U_2 = 0{\cdot}9 \times 10^{-2} \mu b^2 d' \left(\ln\frac{d'}{b_0}\right)^{1/2}, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where

$$d' = \frac{\mu b^2}{36\pi(1-\nu)} \frac{1}{\gamma + n\sigma b/\sqrt{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

is the width of the barrier, ν is Poisson's ratio and γ the stacking fault energy.

Now it is reasonable to assume that the expression

$$X = \left(\ln\frac{h}{b_0}\right)\left(\ln\frac{d'}{b_0} - 1\right)$$

is positive and not much smaller than unity. Then, using (6), (7), (8) and (9), one can see that *the constriction energy U_2 is always negligible* compared with U_1. The activation energy is therefore given by U_1. According to (6), it should be roughly proportional to $(n\sigma)^{-1}$, thus to σ^{-2} ; for, from § 4, n is proportional to $\sigma = \sigma_1 + \sigma_1'$:

$$n = \frac{2L^2\sigma}{bD\mu} \simeq 4 \times 10^6 \frac{\sigma}{\mu}. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

† The loop of recombined barrier will actually soon split again under the applied stress. But the new barrier, having left the (111) glide plane of the piled up group, will no longer retain its dislocations.

For tractions at constant strain rate, *the parabolic hardening should start* at a constant value of U_1/kT, and therefore *at a stress* σ_C *such that* $\sigma_C^2 T$ *is a constant.* Table 1, deduced from Blewitt's measurements on copper 1955), shows that this is approximately the case.

Table 1. Relation between Temperature T of Traction and the Stress σ_C at which Parabolic Hardening occurs (from Blewitt 1955, on copper)

T°K	σ_C dynes cm$^{-2}\times10^{-8}$	$\sigma_C^2 T$ c.g.s. $\times10^{-19}$
4·2	>20	>1·7
78	6·5	3·3
200	4	3·2
300	3	2·7

The value observed for $\sigma_C^2 T$ in these measurements agrees at least roughly with the preceding equations. These give

$$U \simeq U_1 = \frac{(\ln h/b_0)(\ln d'/b_0 - 1)bD}{72\pi^2 K^2 L^2}\left(\frac{\mu}{\sigma}\right)^2 \mu b^3. \quad . \quad . \quad (11)$$

Thus, for copper and with the values $L \simeq 50D \simeq 10^{-3}$ cm of § 4, U/kT is of the order of $3\times10^{19}X/\sigma^2 T$. One can expect for copper values of $X=(\ln h/b_0)(\ln d'/b_0-1)$ between 1 and 5. With $\sigma_C^2 T \simeq 3\times10^{19}$ c.g.s. from table 1, this gives a Boltzmann factor $\exp(-U/kT)$ somewhat too large, but of the right order of magnitude.

Finally eqns. (9) and (11) show that U does *not* depend very much on the stacking fault energy γ if $n\sigma b \gg \gamma$, thus, from (10), for

$$\sigma^2 \gg D\mu\gamma/2L^2 \simeq 10\,\mu\gamma \text{ cm}^{-1}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

It is easy to see that this condition is fulfilled over the range of stresses of the linear stage BC, except at its very beginning. Thus for copper ($\gamma \simeq 40$ erg cm^{-2}) and aluminium ($\gamma \simeq 200$ erg cm^{-2}), (12) gives $\sigma \gg 0{\cdot}15$ and $0{\cdot}25\times10^8$ dynes cm^{-2} respectively; these stresses are somewhat smaller than that σ_B at the beginning of the linear stage (fig. 1).

Table 2. Values Observed for $\sigma_C^2 T(\mu b)^3$ for Various Metals

Metal	T°K	σ_C dynes cm$^{-2}\times10^{-8}$	$[\sigma_C^2 T/(\mu b)^3]$ c.g.s. $\times10^{-7}$
Al	90	4	2·4
	300	≃0·2	≃0·1
Ni	300	>7	>2·2
Cu	300	3	1·9
Ag	300	2	2·5
Au	300	1	0·6

According to (11), *the values of* U *and of* $\sigma_C^2 T$ *should therefore vary as* $(\mu b)^3$ *from metal to metal, except at temperatures* T *high enough for the*

linear stage BC *to be short* ($\sigma_C \simeq \sigma_B$). Table 2, deduced from measurements by Sachs and Weerts (1930), Staubwasser (1954), Kocks (1954), Blewitt (1955), Diehl, Mader and Seeger (1955), shows indeed values of $\sigma_C^2 T/(\mu b)^3$ of the same order of magnitude for various metals, except for room temperature measurements of gold and aluminium, where deviations are expected because the linear stage is short. The value of μ used in the table is that observed for polycrystalline material.

Two other processes may be more favourable in some extreme cases (cf. Stroh 1955):

1. For metals with large stacking fault energies and at low stresses, that is at high temperature, the barrier of fig. 4(*b*) could recombine over a certain length and slip as a whole [110] dislocation, with a (100) slip plane. This could explain the (100) active slip planes observed in aluminium at high temperature (Schmid and Boas 1936).

2. For metals with low stacking fault energies and under high stresses, the free half dislocations of the barrier in fig. 4(*b*) may be driven away from the barrier, thus extending the stacking fault and giving rise eventually to twinning. This could provide a mechanism for the formation of numerous stacking faults (or thin twins ?) in some alloys with low stacking fault energy (Barrett 1952), and the deformation twins apparently observed in copper after fracture (Kuhlmann-Wilsdorf and Wilsdorf 1953) or by traction at very low temperature (Blewitt 1955).

§ 7. Parabolic Hardening

We do not attempt a theory of the parabolic hardening CD (fig. 1), and restrict ourselves to the following remarks.

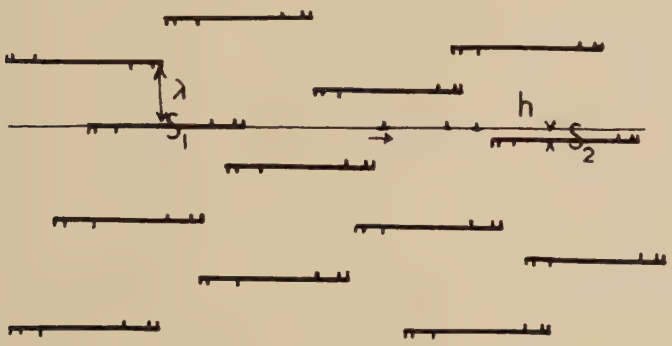
1. There is an obvious relation between the traction experiments discussed in this paper and the yield points observed in ' strain recovery ' by Cottrell and Stokes (1955). These authors strained a crystal in the parabolic region at a temperature T_2 after a prior deformation at a lower temperature T_1. The stress σ_2 required for the deformation at T_2 was lower than the stress σ_1 at the end of the prior deformation at T_1, and dropped further during the deformation, giving a stress–strain curve with an initial yield point.

This behaviour could have been expected : the piled up groups formed under a stress σ_1 at temperature T_1 start to break down, at a higher temperature T_2, under a lower applied stress σ_2. A still lower stress is needed to develop the loops, once they have broken through the barriers, hence the yield point.

2. Once dislocations begin to break through a Cottrell barrier, the loops which have escaped are too big to have much chance of forming a continuous Cottrell barrier again. We can therefore expect them to extend through a large part of the crystal. Their back stress on their emitting sources being negligible, these are expected to go on emitting new loops. Finally these big loops, when gliding in a slip plane near

enough to another piled up group, may help it to break through its barrier and develop into another big slip band (fig. 6). This would

Fig. 6



Production of multiple slip bands in the parabolic stage.

explain the *few*, *big*, *extended* and *multiple* slip bands characteristic of the parabolic stage (Diehl, Mader and Seeger 1955). For slip lines extending over about $\mathscr{L} \simeq 1$ mm, we expect a minimum distance h between the components of a multiple slip band of the order of

$$h \simeq \frac{1}{N_S \mathscr{L}^2} \simeq \frac{12 l^3}{\mathscr{L}^2},$$

that is about 100Å for a Frank–Read net of size $l \simeq 10^{-3}$ cm. This agrees with the distances observed in aluminium (Heidenreich and Shockley 1948, Diehl, Mader and Seeger 1955). It is to this stage that the hardening mechanism by cutting through the forest of screw dislocations discussed by Seeger (1955) can apply.

3. From eqn. (10), the number n of dislocations in the piled up groups for monocrystals* cannot be larger than $4 \times 10^6 \sigma_C/\mu$, if σ_C is the critical stress where parabolic hardening starts at the temperature considered. For aluminium and copper at room temperature, the values of σ_C given in table 2 correspond respectively to $n \simeq 250$ and 2500.

Now, in recovery and high temperature creep, the stress dependence of the activation energy may be related to the value of n, if these processes are regulated by the climb of dislocations in front of piled up groups (Mott 1953). The energy of activation for recovery and steady state creep can then be of the form†

$$U' = \text{const.} - n\sigma b^3, \qquad \ldots\ldots \quad (13)$$

where $n\sigma b^3$ is the work done by the stress $n\sigma$ in front of a piled up group

---

* In polycrystals, larger values of n can of course occur in groups piled up in front of grain boundaries.

† This formula assumes $n\sigma b^3 \geqslant kT$. For $\sigma \geqslant 10^{-4}\mu$ and as $\mu b^3 \simeq 5$ ev in aluminium, this gives $n \geqslant 40$ at room temperature.

when a jog of length b climbs over a distance b. Values $n \simeq 100$ to 200 in aluminium can indeed be deduced from recovery measurements on monocrystals by Kuhlmann, Masing and Raffelsieper (1949, 1950); and also from an analysis of β (transient) creep measured by Wyatt (1953), if one assumes with Mott (1953) an activation energy $\frac{1}{3}U'$ for that type of creep. These last measurements were made on polycrystals with large grains. It would be interesting to repeat them on monocrystals and with other face-centred cubic metals, such as copper and nickel.

As, according to (10), n is proportional to σ, the activation energy U' should vary linearly with σ^2, *not* σ. The measurements made on recovery and creep are, however, not accurate enough to decide which law is observed.

The author wishes to thank Professor N. F. Mott for his constant help and Drs. A. H. Cottrell, B. Jaoul and A. N. Stroh for many enlightening discussions.

APPENDIX A

Formation of Cottrell Barriers in Face-centred Cubic Crystals

It seems reasonable to assume that a Cottrell barrier can be formed readily only if it is stable and if the two component dislocations D_1, D_2 attract each other at large distances.

We shall take the two dislocations parallel to the interaction of their glide planes. And we shall assume their energy of interaction proportional to the scalar product of their Burgers vectors and to the logarithm of their distance. We neglect therefore the differences between screw and edge dislocations, and also the angular dependence of their interaction.

We shall then show that *a screw dislocation cannot form a Cottrell barrier readily*; and that an edge dislocation belonging to a given ' half ' slip system (i.e. a system with a Burgers vector of given *sign*) can form readily a barrier with a given orientation only with edge dislocations belonging to another definite half slip system. *The density of* S′ *sources is therefore equal to that of* S *sources*, as stated in § 3.

A.1. *Short Range Interactions*

Let $(1\bar{1}1)$ and $(\bar{1}11)$ be the two glide planes. By symmetry, it is sufficient to consider for D_1 a screw* [330] and an edge [033] directions. We consider for D_2 the six possible directions of the Burgers vector.

D_1 and D_2 are split into half dislocations X, Y separated by ribbons of stacking faults. The splitting can occur in two ways, leading to stacking faults of different natures (of the type ABC/BCAB and ABC/B/ABC) and half dislocations with different core energies (Frank and Nicholas 1953). It is reasonable to assume the difference of their

* Burgers vectors are given here in units of $b/3\sqrt{2}$.

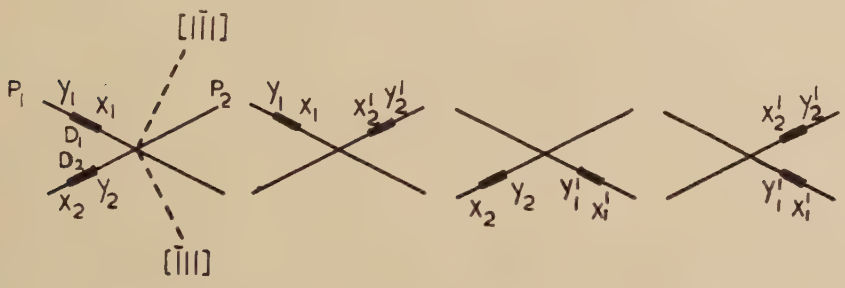
energies large enough for one kind of splitting to occur much more frequently than the other. This splitting, obtained following Thompson's rule (1953), is given in table A 1.

Table A 1. Decomposition into Half Dislocations X, Y

	330	033	330	303	$03\bar{3}$	$\bar{3}30$	$\bar{3}0\bar{3}$	$0\bar{3}3$
X	121	121	211	211	$\bar{1}1\bar{2}$	$\bar{1}2\bar{1}$	$\bar{1}1\bar{2}$	$\bar{1}2\bar{1}$
Y	$21\bar{1}$	$\bar{1}12$	$12\bar{1}$	$1\bar{1}2$	$12\bar{1}$	$2\bar{1}\bar{1}$	$2\bar{1}\bar{1}$	$1\bar{1}2$

The split dislocations can have one of the four dispositions pictured in fig. A 1. The interaction *between the two nearest half dislocations* becomes preponderant when D_1 and D_2 are near to each other. It must therefore be considered first.

Fig. A1



Four possible cases for the interaction of dislocations D_1 and D_2.

(*a*) If these two half dislocations repel each other (cases denoted r, table A 2), there is *no* barrier formed.

(*b*) If they attract each other, the complex dislocation thus formed may repel the two remaining half dislocations sufficiently to form a *stable barrier*. These cases are denoted in table A 2 by the Burgers vector of the complex formed. They are of course linear combinations of the dislocations of the ' stair-rod ' type introduced by Thompson (1953).

(*c*) If the complex does not make a stable barrier with the two remaining half dislocations, it absorbs them to make an ordinary dislocation, denoted in table A 2 by its Burgers vector. This vector is usually in one of the initial glide planes, in which case the dislocation will split into half dislocations and be able to glide, thus providing *no barrier*. In the case of two screws, there is no dislocation left. Finally the combination of [30$\bar{3}$] with [033] leads to dislocations with Burgers

Table A 2. Short Range Interaction between two Dislocations

	$\bar{1}11$ / $1\bar{1}1$	Screw		Edge				Screw		Edge			
		330		303		$03\bar{3}$		$\bar{3}30$		$\bar{3}03$		$0\bar{3}3$	
		X_2Y_2	$X'_2Y'_2$	X_2Y_2	$X'_2Y'_2$	X_2Y_2	$X'_2Y'_2$	X_2Y_2	$X'_2Y'_2$	X_2Y_2	$X'_2Y'_2$	X_2Y_2	$X'_2Y'_2$
Screw	330 X_1Y_1	r	r	r	r	r	$03\bar{1}$	0	0	$03\bar{3}$	$03\bar{3}$	r	303
	330 $X'_1Y'_1$	r	r	301	r	r	r	0	0	$03\bar{3}$	r	303	303
Edge	033 X_1Y_1	r	r	r	r	r	$03\bar{1}$	$\bar{3}03$	$\bar{3}03$	$\bar{1}10$	$\bar{3}30$	r	002
	033 $X'_1Y'_1$	031	r	r	r	031	$\bar{2}20$	$\bar{3}03$	r	$\bar{3}30$	$\bar{3}30$	r	r

vector not in $(1\bar{1}1)$ nor $(\bar{1}11)$; these will split immediately into the more stable complex $[\bar{1}10]$ and two half dislocations, formed directly in one of their dispositions (table A 2).

A.2. *Long Range Interaction*

The splitting of the dislocations may be neglected when they are at large distances. Depending on the angle between their total Burgers vector, they repel each other (R, table A 3), attract each other (A) or do not interact (N).

Table A 3. Long Range Interaction between two Dislocations

$1\bar{1}1$ \ $\bar{1}11$	330	303	$03\bar{3}$	$\bar{3}30$	$30\bar{3}$	$0\bar{3}3$
330	R	R	R	A	A	A
033	R	R	N	A	A	N

The following conclusions may be deduced from tables A 2 and A 3.

1. *Barriers cannot be formed with two screw dislocations.*

2. A screw and an edge dislocation can form a (not very stable) barrier by short range interaction. This is however very unlikely to occur, for the dislocations *repel each other at large distances.*

3. With a given *edge* dislocation ([033] for instance), half of the possible *edge* dislocations in the corresponding intersecting glide plane can form a barrier of a given direction (here [110]) by short range interaction ; and these edge dislocations belong to three ' half ' slip systems (i.e. systems with a Burgers vector of given sign). But only a quarter, belonging to *one half slip system* (here $[\bar{3}03]$) are attracted at long distance by the given edge dislocation, and are thus likely to form a barrier. This barrier is of the type studied by Cottrell (1952). It has the smallest possible Burgers vector, thus is the most stable.

References

Barrett, C. S., 1952, *Imperfections in Nearly Perfect Crystals* (New York: Wiley).

Blewitt, T. H., 1953, *Phys. Rev.*, **91**, 1115 ; 1955, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society).

Blewitt, T. H., and Koehler, J. S., 1950, *Pittsburgh Conference* (Washington : Carnegie Institute of Technology and Office of Naval Research).

Cottrell, A. H., 1953, *Dislocations and Plastic Flow in Crystals* (Oxford : Clarendon Press).

Cottrell, A. H., and Stokes, R. J., 1955, *Phil. Mag.* (to be published).

Diehl, J., Mader, S., and Seeger, A., 1955, to be published.

Frank, F. C., and Nicholas, J. F., 1953, *Phil. Mag.*, **44**, 1213.

Friedel, J., Boulanger, C., and Crussard, C. 1955, *Acta Met.* (to be published).

Heidenreich, R. D., and Shockley, W., 1948, *Bristol Conference* (London : Physical Society).
Jacquet, P. A., 1954, *Acta Met.*, **2**, 742.
Jaoul, B., 1955, *Comptes Rendus*, **240**, 2532; **241**, ; 161.
Kocks, U., 1954, *Stuttgart Conference*. Unpublished.
Kuhlmann, D., Masing, G., and Raffelsieper, J., 1949, *Z. Metallk.*, **40**, 241 ; 1950, *Ibid.*, **41**, 65.
Kuhlmann–Wilsdorf, D., van der Merwe, J. H., and Wilsdorf, H., 1952, *Phil. Mag.*, **43**, 632.
Kuhlmann–Wilsdorf, D., and Wilsdorf, H., 1953, *Acta Met.*, **1**, 394.
Leibfried, G., 1951, *Z. Phys.*, **130**, 214.
Masima, M., and Sachs, G., 1928, *Z. Phys.*, **50**, 161.
Mott, N. F., 1953, *Phil. Mag.*, **44**, 742, 1151.
Röhm, F., and Kochendörfer, A., 1948, *Z. Naturf.*, **3a**, 648.
Rosi, F. D., 1954, *J. Met.*, **6**, 1009.
Sachs, G. and Weerts, J., 1930, *Z. Phys.* **62**, 473 ; 1931, *Ibid*, **67**, 507.
Schmid, E., and Boas, W., 1936, *Kristallplastizität* (Berlin).
Seeger, A., 1954 a, *Phil. Mag.*, **45**, 771 ; 1954 b, *Z. Naturf.* **9a**, 758. 856, 870 ; 1955 a, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society) ; 1955 b, *Phil. Mag.* (to be published).
Staubwasser, W., 1954, *Dissertation*, Göttingen.
Stroh, A. N., 1953, *Proc. Phys. Soc.* B, **66**, 1 ; 1954, *Ibid*, B, **67**, 427 ; 1955, (to be published).
Talbot, J., de Beaulieu C., and Chaudron, G., 1953, *Compte Rendus*, **236**, 818.
Taylor, G. I., 1934, *Proc. Roy. Soc.* A, **145**, 362, 388.
Thompson, N., 1953, *Proc. Phys. Soc.* B, **66**, 481.
von Göler, and Sachs, G., 1929, *Z. Phys.*, **55**, 581.
Wyatt, O. H., 1953, *Proc. Phys. Soc.* B, **66**, 459.

CXXXI. *Effect of Temperature on the Flow Stress of Work-Hardened Copper Crystals*

By M. A. Adams
Department of Physical Metallurgy, University of Birmingham

and A. H. Cottrell, F.R.S.
Atomic Energy Research Establishment, Harwell; previously Department of Physical Metallurgy, University of Birmingham*

[Received August 8, 1955]

Summary

Changes in the flow stress of copper crystals due to changes in the temperature of deformation have been measured over a range of strains and temperatures. After correction for the effect of temperature on elastic constants the flow stress proves to be insensitive to temperature from 90°K to 180°K, falls rapidly with increasing temperature from 180°K to 250°K, and is virtually constant from 250°K to 473°K (the highest temperature examined). The change of flow stress with temperature is closely proportional to the flow stress itself. These results are discussed briefly in terms of recent ideas of dislocation processes sensitive to temperature. The effect of increasing the temperature of deformation is to produce a yield drop, similar to those observed in aluminium crystals during work softening.

§ 1. Introduction

It has been realized in recent years that in addition to the elastic interactions between parallel dislocations (Taylor 1934) several other processes involving dislocations may contribute to work hardening (Mott 1952, 1953); for example the coalescence of dislocations to form sessiles (Lomer 1951, Cottrell 1952 a), the formation of jogs on intersecting dislocation lines (Heidenreich and Shockley 1948, Cottrell 1952 b, Seeger 1954 a and b), and the creation of vacancies and interstitials at jogs (Mott 1951, Seitz 1952, van Bueren 1953, Seeger 1955). It is therefore important to develop experimental methods for separating and identifying the respective contributions of these processes to work hardening. One possible method of attack, which was chosen for the present experiments, is to determine the change of flow stress with temperature. This ought to provide

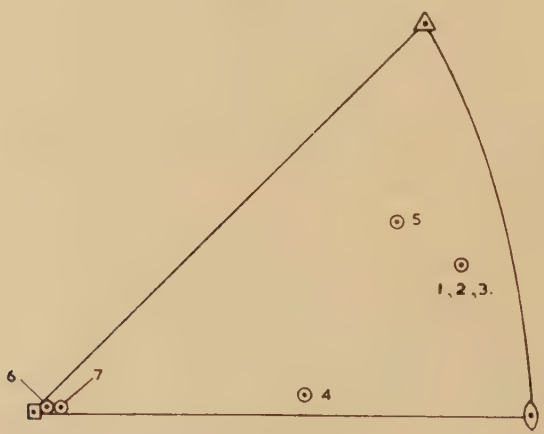
* Communicated by the Authors.

information about the contributions of the processes according to their sensitivity to temperature. In these experiments single crystals of copper were used; the results of similar experiments on aluminium crystals have recently been described (Stokes and Cottrell 1954, Cottrell and Stokes 1955).

§ 2. Experimental Method

Wires of 1·5 mm diameter were prepared by swaging rods of both spectroscopically pure (99·999%) and oxygen-free high-conductivity (99·99%) copper. After pickling and coating with colloidal graphite, these wires were grown into single crystals 17 cm long by the fusion method of Andrade and Roscoe (1937), being supported horizontally during growth in loosely fitting and carefully cleaned silica quills. The crystals were cleaned in dilute nitric acid and cut into 6 cm lengths with a fine oxy-acetylene flame, water cooling being used to protect sections near the cut. Small iron loops which enclosed 1 cm lengths at the ends of the crystals were then attached by silver solder, thus leaving a working length of 4 cm. The completed specimens were electropolished in a phosphoric acid cell and their orientations determined by the Laue back-reflection method (fig. 1).

Fig. 1



Initial Orientations of the Specimens. Numbers 4 and 5 were made from oxygen-free high-conductivity copper; the remainder from spectroscopically pure copper.

Tensile experiments were made with a hard-beam machine at a strain rate of 4×10^{-5} per sec. A small load was always maintained on each specimen, between successive experiments, to preserve the alignment of the hooks and loops through which the load was applied. To maintain a selected temperature the mounted specimen was enclosed in a thermal bath attached to the machine. A standard time of 15 min was allowed in each bath before the experiment was begun and about 2 min was

needed to replace one bath by another. The following temperatures were used :—

Temperature °K	Bath	Fixed Point
90	liquid oxygen	boiling point
196	solid carbon dioxide and acetone	sublimation point
233	solid and liquid calcium chloride solution	freezing point
273	water and ice	freezing point
373	silicone oil	controlled to ± 1°K
473	silicone oil	controlled to ± 1°K

The general method of making experiments was to strain each crystal by repeated increments, each about 0·01, applied alternately at two temperatures. The load and elongation were recorded continuously and the effect of temperature on the flow stress was determined by comparing the flow stress of the specimen at the end of one increment with that at the beginning of the next increment.

§ 3. Results

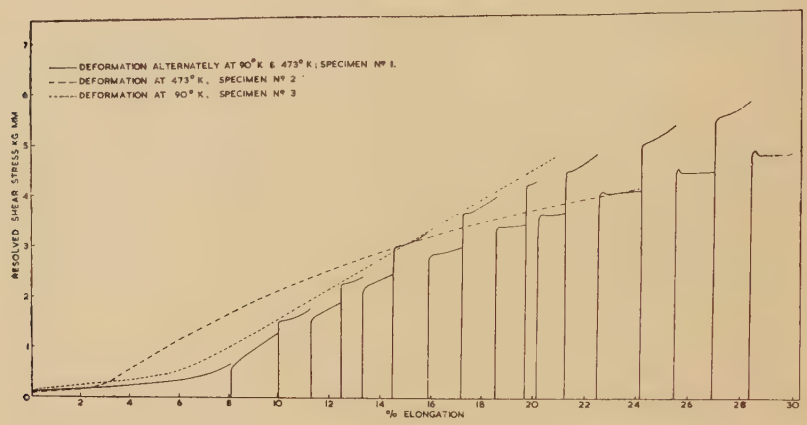
Two series of stress-elongation curves are shown in figs. 2 and 3, obtained from crystals of spectroscopically pure and oxygen-free high-conductivity copper, respectively, cycled between 90°K and 473°K. Figure 2 in addition shows two isothermal curves obtained from crystals of the same orientation as that used in the cycling experiments, taken at 90°K and 473°K respectively.

These isothermal curves show that at small strains more work hardening occurs at 473°K than at 90°K, as a result of the shorter period of easy glide at 473°K. At large strains, however, the crystal cannot maintain the same high level of flow stress at 473°K as it can at 90°K ; the rate of work-hardening falls off at 473°K but not at 90°K. This latter effect, which has also been observed by Blewitt, Coltman and Redman (1955), seems to be related to the yield phenomenon which appears in the strongly worked crystal on straining at 473°K after a preliminary straining at 90°K. In a study of similar yield drops in aluminium crystals, Cottrell and Stokes (1955) showed that a cold-worked crystal softened extremely rapidly when strained again at a higher temperature than that of working. They suggested that this work-softening effect was caused by the collapse, due to the joint action of stress and thermal agitation, of sessile dislocations created during the previous straining at the low temperature. This thermally activated collapse of sessile dislocations during plastic flow has also been suggested independently by Friedel (J. Friedel, private communication and *Phil. Mag.*, 1955) to explain the observations of Blewitt, Coltman and Redman on the effect of temperature on the work-hardening of copper crystals at large strains.

From the point of view of measuring the reversible change of flow stress with change in the temperature of straining, the yield drops which occur

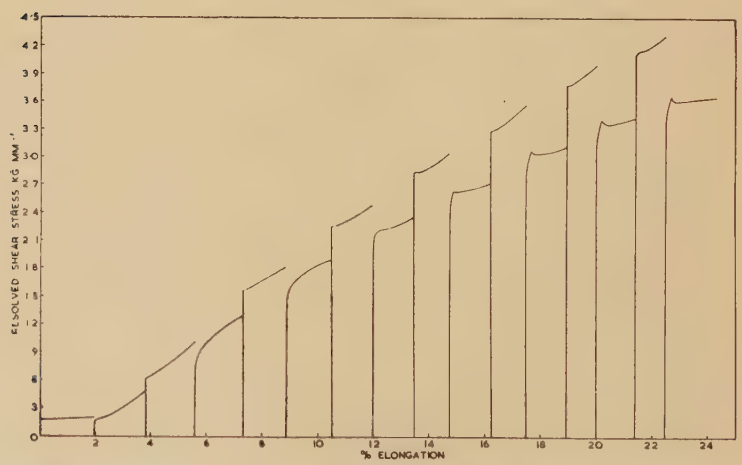
after changing from low to high temperatures are an unwelcome complication since they add an irreversible change of flow stress to the reversible one. Changes from high to low temperature on the other hand do not suffer from this effect. Accordingly, all the measurements of changes of flow stress reported below have been obtained from transitions of this second type.

Fig. 2



Stress-elongation curves on crystals of spectroscopically pure copper.

Fig. 3



Stress-elongation curves on a crystal of oxygen-free high-conductivity copper (specimen number 4) deformed alternately at 90°K and 473°K.

Changes in flow stress produced by changes from 473°K to 90°K made at various stages during the elongation of the crystal are given in the table, taken from the results in figs. 2 and 3. An important feature is that the ratio of flow stresses at these two temperatures settles down after an

elongation of 10–15% to a constant value that is practically the same for both grades of copper. This observation, which has also been made on aluminium (Cottrell and Stokes 1955), encourages the view the ratio of

The Ratio of Flow Stress at 473°K to that at 90°K, Measured on Copper Crystals by Transitions from 473°K to 90°K at Various Elongations

Specimen Number 1

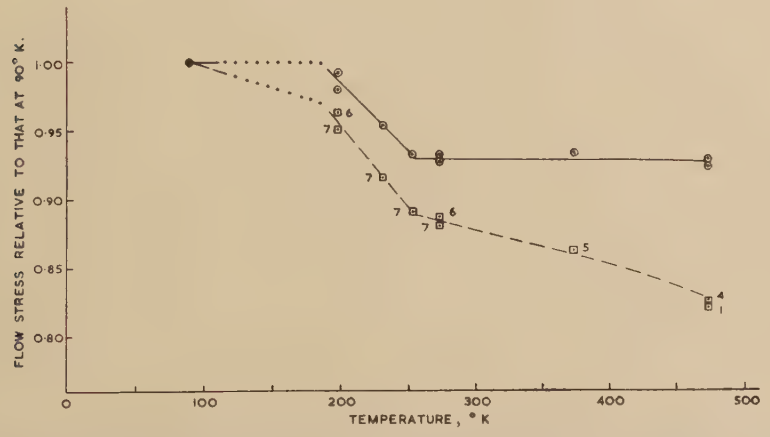
Per cent elongation :	10·0	12·5	14·5	17·2	19·7	21·2	24·2	27·0
Ratio of flow stress :	0·858	0·854	0·824	0·82	0·822	0·826	0·820	0·821

Specimen Number 4 :

Per cent elongation :	3·8	7·3	10·5	13·5	16·3	1·90	21·4
Ratio of flow stress :	0·783	0·819	0·837	0·832	0·828	0·823	0·834

flow stresses, measured in this way, may provide a clue to the nature of the elementary events involving individual dislocations ; the absolute magnitude of the flow stress, on the other hand, depends on the number of such events as well as on their nature. The ratios of flow stresses obtained from all the subsequent experiments were measured in this constant range.

Fig. 4



The reversible change of flow stress with temperature for work-hardened copper crystals, before (lower curve) and after (upper curve) correction for the change of elastic constant with temperature. The numbers of the specimens are shown.

Figure 4 summarizes the results of these experiments and gives the flow stress at various temperatures relative to that at 90°K. All values except those at 195°K were obtained by direct transitions from the temperature concerned to 90°K. To avoid contact between liquid oxygen and acetone, the values for 196°K were obtained indirectly, by double transitions from 273°K to 196°K and from 273°K to 90°K. The broken curve in fig. 4 shows the ratio of flow stresses directly. Since the elastic constants of the metal vary with temperature it is important to examine the variation of

σ/μ with temperature, where σ and μ are the flow stress and shear modulus measured at each temperature. In the ' corrected ' flow stress curve (full line) of fig. 4 the values of Köster (1948) for the variation of Young's modulus with temperature have been used, since this variation is very similar to that of shear modulus.

Correcting for the effect of elastic constants brings out the remarkable fact that the residual sensivity of the flow stress to temperature is confined almost entirely to a ' shelf ' from 180°K to 250°K across which the flow stress falls from near unity to 0·93. Outside this range of temperature there is very little variation with temperature; this constancy is particularly striking over the range from 250°K to 470°K where the corrected flow stress remains entirely inside the limits 0·924 and 0·934.

§ 4. Discussion

The main processes suggested so far that may provide large reversible changes of (corrected) flow stress with temperature in a metal such as copper at low temperatures are (*a*) the intersections of dislocation lines gliding through a ' forest ' of dislocations (Cottrell 1952 b), (*b*) the thermal creation of point defects at jogs on moving dislocations with screw components (Seeger 1955), and (*c*) the thermally activated glide of jogs on edge dislocations (Maddin and Cottrell 1955). A detailed interpretation of fig. 4 in terms of the first two of these processes has already been given by Seeger (1955). His theory, which we shall now outline briefly, depends on the differences in behaviour of edge and screw dislocations and on the principle that at any given temperature and rate of straining, the strain is produced mainly by the type of dislocation that is easier to move. At low temperatures (i.e. below 180°K) the edge dislocations move more easily than the screws because the latter must create vacancies (or interstitials) at jogs whenever they move. The moving edge dislocations cut their way through the forest and this leads to a small thermal dependence of flow stress in this low-temperature range. Above this range it becomes easier to move screw dislocations than edges because thermal fluctuations can play an increasingly important part in creating vacancies at jogs in screw dislocations: the steep fall of flow stress at 200°K is attributed to this effect. Above a certain temperature the generation of vacancies at jogs on screw dislocations occurs so easily that it is no longer an appreciable source of resistance to the motion of screws. Accordingly, the strength remaining in this high-temperature range is due mainly to long-range elastic interactions of dislocations. Since these elastic interactions depend on temperature only through the elastic constants the (corrected) flow stress is constant at these temperatures.

Seeger's theory attributes the flow strength of copper in various temperature ranges to several distinct dislocation processes. Looked at from this point of view the results in the table are intriguing since they

imply that over a wide range of cold-worked states the change of flow stress with temperature is closely proportional to the flow stress itself. The theory thus requires a proportionality to be maintained during work-hardening between the density of forest dislocations, the density of jogs on screw dislocations, and the density of internal stresses in the metal. Such a proportionality might be expected if the various distributions of dislocations produced in the present series of experiments differed only in scale, and not in pattern. Clearly, an obvious extension of these experiments (which it is hoped soon to make) is to measure the flow stress as a function of temperature in crystals containing radically different patterns of dislocations.

Acknowledgments

The authors are grateful to Professor N. F. Mott, F.R.S. and Dr. R. J. Stokes for useful discussions during the course of this work, and to Dr. A. Seeger for showing them his manuscript before its publication.

References

Andrade, E. N. da C., and Roscoe, R., 1937, *Proc Phys. Soc.*, **39**, 152.
Blewitt, T. H., Coltman, R. R., and Redman, J. K., 1955, *Report on Defects in Crystalline Solids* (London : Physical Society), p. 369.
Cottrell, A. H., 1952 a, *Phil. Mag.*, **43**, 645 ; 1952 b, *J. Mech. Phys. Solids*, **1**, 53.
Cottrell, A. H., and Stokes, R. J., 1955, *Proc. Roy. Soc.*, in press.
Friedel, J., 1955, *Phil. Mag.* (this number).
Heidenreich, R. D. and Shockley, W., 1948, *Report on Strength of Solids* (London : Physical Society), p. 57.
Köster, W., 1948, *Z. Metallk.*, **39**, 1.
Lomer, W. M., 1951, *Phil. Mag.*, **42**, 1327.
Maddin, R., and Cottrell, A. H., 1955, *Phil. Mag.*, **46**, 735.
Mott, N. F., 1951, *Proc. Phys. Soc.* B, **64**, 729 ; 1952, *Phil. Mag.*, **43**, 1151 ; 1953, *Ibid.*, **44**, 842.
Seeger, A., 1954 a, *Phil. Mag.*, **45**, 771 ; 1954 b, *Z. Naturforsh.*, **9**, 758 ; 1955, *Phil. Mag.* (to be published).
Seitz, F., 1952, *Advances in Physics*, **1**, 43.
Stokes, R. J., and Cottrell, A. H., 1954, *Acta Met.*, **2**, 341.
Taylor, G. I., 1934, *Proc. Roy. Soc.* A, **145**, 362.
van Bueren, H. G., 1953, *Acta Met.*, **1**, 464.

CXXXII. *The Generation of Lattice Defects by Moving Dislocations, and its Application to the Temperature Dependence of the Flow-Stress of F.C.C. Crystals*

By Alfred Seeger

Max-Planck-Institut für Metallforschung, and Institut für theoretische und angewandte Physik der Technischen Hochschule Stuttgart, Stuttgart, Germany*

[Received April 4, 1955; revised July 1, 1955]

Summary

The present paper treats in some detail the various mechanisms by which lattice defects are formed by moving dislocations. Particular attention is paid to the role of thermal energy in these processes and to the special features of extended dislocations in metals with low stacking fault energies. The results are applied to the temperature dependence of flow stress of such metals and are found to explain the temperature dependence observed in Cu quantitatively. Cu is known to have a low stacking fault energy.

It is possible to deduce details about the mechanism of plastic deformation, in particular of the role played by the edge and the screw parts of dislocation rings under various conditions. The predictions of the theory on the nature of the lattice defects causing the increase of electrical resistance during cold work are shown to be borne out by recovery experiments.

§ 1. Introduction

It has been known for some time that moving dislocations may generate lattice defects (for a review see Seitz 1952). If we confine ourselves—as we shall do throughout this paper—to temperatures below that of self-diffusion so that climbing of dislocations is unimportant, the accepted mechanism is that a jog in a moving screw dislocation will leave behind it a row either of vacancies or of interstitial atoms. Very little attention has so far been paid to the finer details of these processes and in particular to the influence of thermal energy on the mode of production of vacancies and other lattice defects.

In § 1 we undertake a study of the various processes that lead to a generation of lattice defects by gliding dislocations. We shall investigate in particular the temperature dependence of these processes, giving special consideration to the conditions under which vacancies are generated thermally by jogs in extended dislocations.† Most of the discussions in the present paper refer to *elementary* jogs (i.e. jogs connecting two pieces of a dislocation line in two *neighbouring* parallel glide planes). Sometimes composite jogs, which are for example formed in dislocations in the h.c.p.

* Communicated by the Author.

† A more detailed description of jogs, their formation and annihilation has been given elsewhere (Seeger 1955 a).

lattice that cut through dislocations with a Burgers vector perpendicular to the basal plane (Seeger 1955 a), are also discussed. Throughout our discussion we shall confine ourselves to unit dislocations in the f.c.c., b.c.c. or h.c.p. structures, thereby avoiding the complications that arise in the case of dislocations with large Burgers vectors.

The application of these results to the temperature dependence of the flow stress in f.c.c. cubic metals is based on the author's earlier discussions of the temperature dependence of plastic properties of metals crystallizing in close-packed structures (Seeger 1954 a, 1955 b). In these papers a detailed account of the temperature dependence of the flow stress of metals with high stacking fault energies was given. The present treatment (§ 3) of metals with low stacking fault energies (in particular noble metals, Ni and Pt) completes the general discussion (Seeger 1955 b) and supplies numerical values for some of the activation energies involved. Very good agreement with experiment is obtained for Cu, the only metal for which suitable experimental results are available. Qualitative agreement is reached for Ni. The implication of this is that Ni must also be considered to have a low stacking fault energy. As in one of the author's earlier papers (Seeger 1954 a) the (not very extensive) experimental results on Ni had been tentatively interpreted on the assumption of a high stacking fault energy we discuss in Appendix A the present evidence on the stacking fault energy of Ni.

The good agreement between theory and experiment for Cu allows us to draw further conclusions on the mechanism of plastic deformation at low temperatures of this metal, which is believed to be representative of the noble metals group. It turns out that at 180°K a transition occurs from a low temperature mechanism in which the edge dislocations move more easily than the screw dislocations to a mechanism in which the screws move more easily than the edges so that the strain rate imposed in a tensile test is maintained mainly by the motion of dislocations with a large screw component. This change in mechanism is accompanied by a change of the nature of the lattice defects created during deformation (e.g. vacancies, pairs or clusters of vacancies, and interstitial atoms). As these lattice defects are known to be responsible for a large portion of the change of electrical resistance during cold work it should be possible to check these predictions by recovery experiments, and the experimental data confirm the theoretical expectations. An annealing stage at about 100°C, which we believe to be due to the migration of isolated vacancies, is pronounced only if the specimens were cold-worked at sufficiently high temperature. In specimens deformed at very low temperatures, in which according to the theory very few isolated vacancies are generated, this annealing stage is found to be weak (§ 4).

§ 2. Athermal and Thermal Generation of Lattice Defects

Thermal activation may or may not be important in the production of lattice defects by moving dislocations. As the effect on plastic properties, e.g. on the temperature dependence of flow stress, will be different for

thermal and athermal generation of lattice defects we start by reviewing the different ways of creating defects by moving dislocations and discussing the role of thermal energy in each of them.

For simplicity we first consider pure screw dislocations with elementary jogs in them We shall later return to the general case that the dislocation lines have both screw and edge components.

If the shear stress τ acting in the glide plane is sufficiently large, both vacancies and interstitials may be generated in strings (fig. 1). τ has to be large enough to overcome the line-tension of the string. The energy necessary to add an interstitial or a vacancy to the end of the string may be written

$$U_1=\alpha_1 Gb^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

(G shear modulus). In typical metals (e.g. Cu) α_1 is of the order of unity for interstitials and about 0·1–0·2 for vacancies. If the distance between neighbouring jogs is l_1, we obtain from considerations of the work done by the shear stress during the generation of a defect (Cottrell 1952)

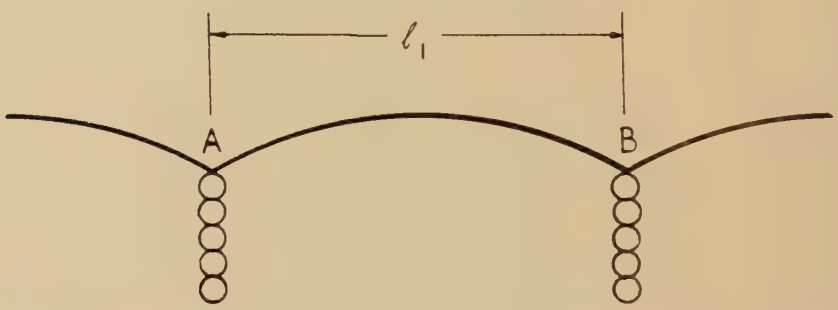
$$\tau=U_1/b^2 l_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2\,a)$$

or, by using (1),

$$\tau=\alpha_1 Gb/l_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2\,b)$$

Equation (3 *b*) is also the equation for the critical stress of the Frank–Read source of length l_1 that extends between the points A and B in fig. 1. In

Fig. 1



Generation of strings of lattice defects by a gliding dislocation of screw character. A and B are the position of the jogs in the dislocation line.

this case the numerical factor α_1 in eqn. (2 *b*) is about unity and will be denoted by α_s.

We can now state the conditions under which (in the absence of thermal activation) defects are continuously added to the ends of the rows of defects shown in fig. 1 :

(i) The shear stress τ must be equal to or slightly larger than the shear stress given by eqn. (2 *a*).

(ii) The shear stress τ must satisfy the inequality

$$\tau \leqslant \alpha_s Gb/l_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

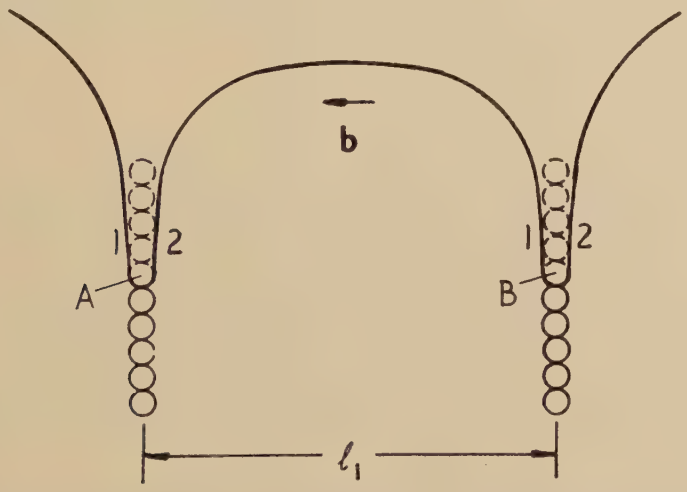
There are many cases in which it is possible to satisfy the conditions (i) and (ii) simultaneously, e.g. the production of vacancies by elementary jogs in f.c.c., b.c.c., and h.c.p. lattices. In the case of composite jogs the situation is more difficult, but it seems safe to assume that in the simple cases just mentioned our conditions can still be fulfilled for jogs composed of two elementary jogs.

The situation is much more critical if the direction of motion of the jogs is such that they will produce interstitial atoms. Since for the generation of one interstitial at a time α_1 is about equal to α_s, the satisfying of our conditions depends on the metal under discussion. If the jogs are composite ones the critical shear stress α_s of the Frank–Read-sources, given by

$$\tau_s = \alpha_s bG/l_1, \qquad (4)$$

will be reached before the shear stress is large enough to add new groups of interstitials to the end of the strings. The dislocation lines between A and B will bulge out, sweep part of the glide plane and unite as shown in fig. 2.

Fig. 2



Generation of a series of lattice defects due to the action of the Frank–Read source of length l_1 with anchor points A and B.

The attraction of the edge dislocations 1 and 2 will certainly be strong enough to annihilate the dislocations, the annihilation products being interstitials. It requires, however, a more detailed investigation to decide whether the local heating during the annihilation process is large enough to allow the interstitials to diffuse away (see Seitz 1952). If the interstitials do not diffuse away the situation is the same as if the interstitials had been created consecutively.

Thermal activation is not important in either of these processes for the production of lattice defects, unless the temperature is high enough to allow the defects to diffuse away from the strings. As pairs and clusters of

vacancies almost certainly have a smaller activation energy of migration than a vacancy, a string of vacancies will presumably break up into groups of vacancies rather than isolated vacancies, either immediately after formation or during an anneal. The essential feature in both cases discussed above is the balancing of the line tensions of the dislocations and the strings of lattice defects by the shear stress τ acting on the dislocations. Thermal energy aids the shear stress only to the extent to which it lowers these line tensions, which are free energies per unit length and depend therefore somewhat on temperature.

If we consider dislocation lines with an increasing ratio of the edge component to the screw component, the distance between neighbouring defects in the strings produced by the forward motion of jogs will increase until we can regard the defects in a string as independent of one another. Then in eqn. (2) U_1 has to be replaced by the energy U_0 necessary for creating isolated defects. We can write

$$U_0 = \alpha_0 G b^3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The remarks on the magnitude of α_1 also apply to α_0.

At temperatures low enough for thermal energy to be unimportant the situation is much the same as with pure screw dislocations. Depending on whether

$$\alpha_0 < \alpha_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6\,a)$$

or

$$\alpha_0 > \alpha_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6\,b)$$

the defects will either be produced consecutively or many at a time through the action of the dislocation source and subsequent annihilation of the dislocation lines.

If condition (6 *a*) is satisfied the position is quite different if thermal energy is sufficient to aid the production of vacancies by elementary jogs. This may be envisaged as occurring in two ways.

(*a*) In creep it has been suggested by Mott (1955) that a jog, momentarily at rest, can move forward one atomic distance with the help of the stress, the vacancy (or interstitial) moving away at the same time. The activation energy for this will be

$$W + U_0 - \tau b^2 l_1$$

where W is the activation energy for the movement of a vacancy, so that in f.c.c. metals $W + U_0$ is that for self-diffusion.

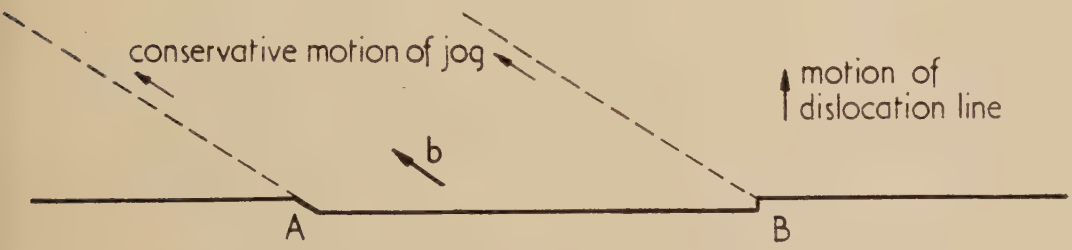
(*b*) In reasonably rapid extension, however, as considered in this paper, we suppose that the jogs in a screw dislocation (which after all are nothing but short edge dislocations) will continuously move up and down the dislocation line under the influence of the applied stress and the rapidly fluctuating internal stresses due to work-hardening. A jog can therefore jump forward while in motion sideways and continue its motion sideways. The hole thus formed does not have to diffuse away to be stable, so the activation energy for this process is

$$U_0 - \tau b^2 l_1.$$

This process is the one considered in this paper and we shall return to it in § 3.

We now discuss in more detail this possibility for the jog to move sidewards, i.e. along the dislocation line. For jogs in a pure edge dislocation, sideward motions of the jogs are always non-conservative and therefore unlikely to occur at very low temperatures. On the other hand, in a pure screw dislocation the sideward motion of a jog is conservative, whereas the forward motion is non-conservative and leads, as discussed above, to the production of lattice defects. This general rule is that a jog in a dislocation line with Burgers vector **b** will move conservatively in the direction of **b**. If **b** is not exactly parallel to the dislocation line (i.e. the dislocation has an edge component), even a dislocation with jogs may glide without the production of lattice defects, the jogs moving conservatively along the dotted lines shown in fig. 3.

Fig. 3



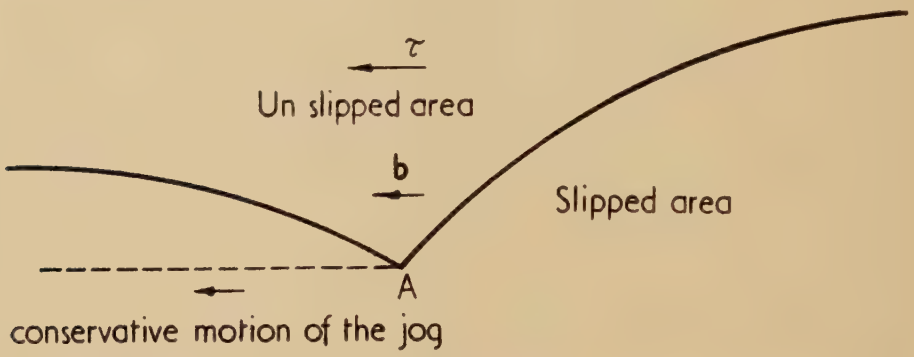
Conservative motions of jogs A and B in a dislocation line.

In a tensile test an unbalanced force due to the applied shear stress and the stress-fields of the surrounding dislocations rearranging themselves during the work-hardening will in general act on a jog in a screw dislocation. If the dislocation itself is held up at some obstacle this force will cause the jog to move along the dislocation line. Figure 4 shows an example in which the area in the glide plane swept out by a dislocation whose mean character is that of a screw can be increased by the movement of a jog. But even if the screw dislocation containing the jog is straight (apart from the jog) work will be done by the shear stress during the motion of the jog. This is so because the force on a dislocation line will in general have a component perpendicular to the glide plane. Since by the movement of the jog the dislocation is lifted from its glide plane to a neighbouring plane this force causes to move the jog along the dislocation line.* In this mechanism jogs of opposite sign move in opposite directions. If a pair of jogs of opposite sign meet each other they will recombine, thereby reducing the number of jogs in screw dislocations. Irrespective of the distance travelled by a screw dislocation through the forest an average distance l_j between dislocations corresponding to a dynamical equilibrium

* It is assumed that the screw dislocation cannot yield by cross slip to the force acting on it. This assumption is satisfied in metals with low stacking fault energies, to which the present considerations will be applied in § 3.

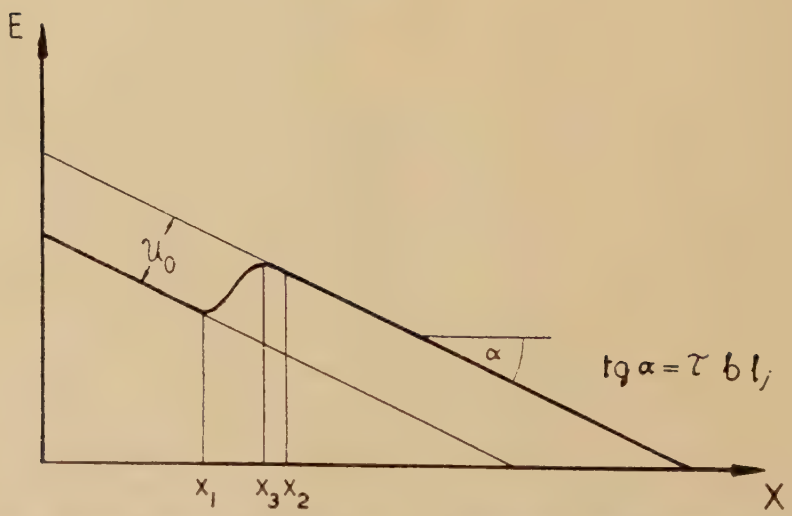
will be established. l_j will depend on such parameters as the density of the forest of dislocations threading the glide plane, the velocities of both dislocation lines and jogs, and the overall strain rate $\dot{\epsilon}$. As the motion of jogs along the screw dislocations is envisaged not to depend on thermal activation, temperature will influence l_j only indirectly through the variation of these parameters.

Fig. 4



The shear stress, tending to increase the slipped area in the glide plane, will cause the jog A to move conservatively to the left.

Fig. 5



Energy versus distance diagram for a dislocation line containing a jog. During the movement from x_1 to x_2 the jog creates a vacancy. The maximum of the potential energy is reached at the position x_3. The applied shear stress is τ.

Since the forces per unit length acting on the jog are of the same order of magnitude as those acting during the glide motion of the dislocations the usual arguments (see e.g. Nabarro 1951) can be applied as to

the velocity of a jog. The velocity attained by a jog under these forces should be of the order one-half of the sound velocity or slightly less. As we shall see in the subsequent discussion, velocities of this magnitude are sufficient to allow the production of isolated vacancies by the motion of jogs.

Figure 5 shows the diagram of energy as a function of distance for a screw dislocation which contains a jog. For the production of a vacancy by a non-conservative movement of the jog we adopt the following model: the vacancy is created during the forward motion of the dislocation line by a distance $x_2-x_1=b$. The maximum of energy during this movement is reached at a position x_3, given by

$$x_3=x_1+3b/4=x_2-b/4. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

With a shear stress τ acting on the dislocation the activation energy for motion in forward direction is

$$U=U_0-3l_jb^2\tau/4. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The activation energy for the jog to jump backwards, thereby recombining with the vacancy, is

$$U=l_jb^2\tau/4. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Equation (9) shows that in order to prevent a substantial fraction of the jogs from recombining with the vacancy τ must not be too small. Otherwise eqn. (8) would have to be amended by a term taking into account this possibility. We shall base our discussion on fig. 9, which applies to Cu. The full line denoted by 3 corresponds to the process under discussion. The lowest stress τ, for which eqn. (8) is used in that diagram, is that at the temperature T_2 and is given by

$$\tau=\frac{1}{4}\tau_a^{(3)}=\frac{1}{3}\frac{U_0}{b^2l_j} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

(for further details of notation see § 3). The frequency of backward jumps is

$$\nu=\nu_0 \exp\left(-\frac{U_0}{12kT_2}\right)$$

which, with $U_0=0{\cdot}86$ ev and $T_2=250°$K, gives

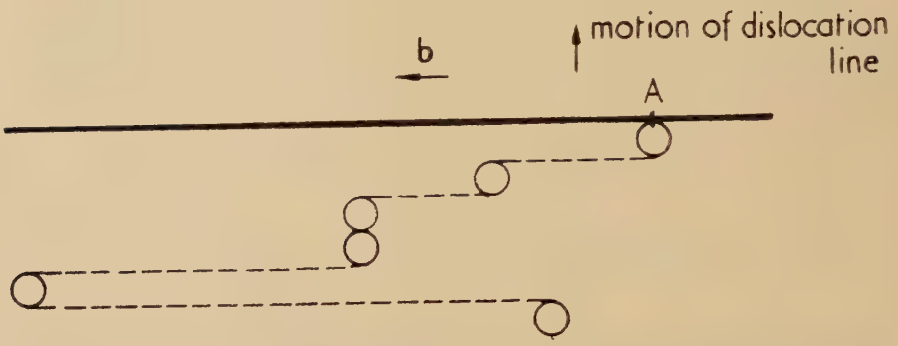
$$\nu=0{\cdot}05\ \nu_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

ν_0 will be of the order of Debye's frequency. Therefore the velocity of the jog along the dislocation must be larger than 0·05 of the sound velocity if the probability of jumping backward and recombining with the vacancy is to be neglected. On the basis of the discussion given above we believe that this condition is satisfied in tensile tests of typical metals.

If the conditions just discussed are fulfilled a jog in a screw dislocation will carry out a zig-zag motion, consisting mainly of a rapid movement along the dislocation line, from time to time interrupted by thermally

activated steps perpendicular to the Burgers vector, in which isolated vacancies are created, and possibly reversals of the direction of motion. This is shown schematically in fig. 6. The jog's motion will end in a recombination with a jog of opposite sign.

Fig. 6



Zig-zag motion of a jog A in a screw dislocation The dotted lines represent the conservative movements of the jog, the circles represent the lattice defects created in non-conservative motions.

If in the process just discussed a complete separation of the defect and the dislocation line occurs the activation energy of the process is given by the energy U_0 of the free defect minus the work done by the applied shear stress. This has been made use of in eqn. (8). If the dislocation involved has exactly screw character it is, however, likely that the defect will remain attached to the dislocation line immediately after its creation. The activation energy will be lowered by the ' binding energy ' U_3 and will be given (the influence of the applied stress τ being neglected) by

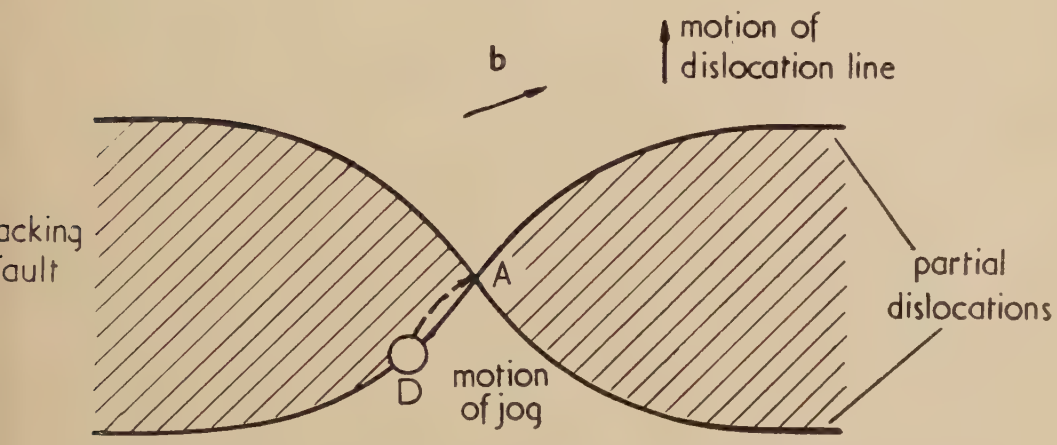
$$U_2 = U_0 - U_3. \qquad . \; . \; . \; . \; . \; . \; . \; . \; (12)$$

Having in mind the application of these considerations to the flow stress of close-packed metals, we consider in some detail the case of a f.c.c. or h.c.p. structure in which there is an appreciable separation of complete dislocations in close-packed glide planes into partial dislocations. At the position of a jog the partial dislocations of the extended dislocation have been brought together, forming a constriction (Stroh 1954, Seeger 1954 c, Schoeck and Seeger 1955). If the jog and therefore also the constriction move away from a defect which has just been generated, the defect will remain attached to one of the partial dislocations as shown in fig. 7.

The energy U_3 by which an interstitial or a vacancy is bound to a dislocation in a f.c.c. metal consists in general of two contributions, namely an interaction energy U_{el} due to the electric interaction between the lattice defect, and an energy term U_{dis} which arises from the elastic interaction of the lattice defect.

An edge dislocation in a metal containing ' free ' electrons will form an electrical line dipole (Landauer 1951, Cottrell, Hunter and Nabarro 1953), the compression side being positively charged and the dilatation side being negatively charged. A lattice defect in a metal such as copper can be looked upon as a negative elementary charge which is screened by the electrons in the Fermi gas (Fumi 1954). Due to the spatial variation of the electric field around a dislocation a vacancy on the compression side of the dislocation will be repelled if sufficiently far away from the centre of the dislocation. This results from the fact that the attraction of the negatively charged centre of the vacancy is more than overbalanced by the repulsion of the positively charged screen. If, however, the vacancy is so close to the dislocation line that the screen overlaps appreciably into the dilated region around an edge dislocation, a net attraction results. At a distance of about 1 Å from the dislocation centre, where a maximum of the elastic interaction is to be expected, the two opposing effects discussed

Fig. 7



A jog (A) in an extended dislocation having created a lattice defect moves away from the defect (D) which remains attached to one of the partial dislocations.

above cancel, so that for the present problem the electrical interaction between dislocations and vacancies may be neglected. It should be pointed out, however, that these results are based on a classical calculation of the interaction between the charge distributions around a vacancy and around a dislocation and that a more refined quantum mechanical treatment might change some of the details. For impurity atoms that do not cause a lattice distortion the elastic interaction can be neglected in comparison with the electric interaction. In such a case the foregoing considerations apply fully, whereas in the case of vacancies in Cu and probably in all noble metals only the elastic interaction needs to be considered.

In f.c.c. structures the lattice distortions due to both interstitials and vacancies show cubic symmetry. We may therefore to a good approximation use the formula (Bilby 1950)

$$U_{\text{dis}}=4G\delta r_0^3 \sin\phi/R \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13$$

to calculate the elastic energy of interaction between a dislocation and a lattice defect whose position with respect to the dislocation line (edge-dislocation of dislocation strength b) is fixed by the polar coordinates R, ϕ. r_0 is the atomic radius, and δ relates r_0 to the apparent radius

$$r=(1+\delta)\, r_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

of the defect. From the Born–Mayer potential proposed by Huntington (1953) for the interaction of the ion cores of Cu^+, a value for δ of 0·0175 is easily deduced (Fumi 1955). The modulus of $b \sin\phi/R$ will be assumed to be equal to unity. A change in this term by a factor of $\frac{1}{2}$ would change the final result only by 5%. For want of better information we use the same numerical values also for Ag, Au, Ni and Pt. The interaction energies U_{dis} are given in the table.

Data referring to the Generation of Lattice Vacancies at Jogs in Extended Dislocations in various f.c.c. Metals

Metal	Cu	Ag	Au	Ni	Pt
Energy U_0 of a free vacancy (ev)	0·90	0·8	0·67	1·6	1·2
Shear modulus G (kg/mm²)	4400	2700	2800	7600	6600
Atomic radius r_0(A)	1·275	1·444	1·442	1·245	1·378
$U_{\text{dis}}=U_3$ (ev)	0·04	0·03(5)	0·03(6)	0·06(3)	0·07(5)
$U_2=U_0-U_3$	0·86	0·7(6)	0·63	1·5	1·1
S^V/k	3·4	(2·5)	1	(4)	(1)
$T_0^{(3)}$ (theoretical, °K)	364	(330)	290	(420)	(530)

Values in brackets less certain than others.
For significance of $T_0^{(3)}$ and comparison with experiment see § 3.

The sources of the energies U_0 of formation of free vacancies used in the table are stated in Appendix B. The activation entropies $S^{(V)}$ for the formation of free vacancies for Cu and Au are deduced from experiment (see § 3), those for Ag, Ni and Pt are estimates.

§ 3. Temperature-Dependence of Flow Stress in Close-Packed Metals

In f.c.c. and h.c.p. metals the main processes which contribute to the temperature-dependence of the flow stress below the temperature of self-diffusion are the intersection of dislocation lines and the production of

lattice defects with the aid of thermal energy. The mechanism of the latter process has been discussed in § 2. The Peierls force which would also give rise to a temperature dependence of the flow stress is thought to be unimportant for f.c.c. and h.c.p. metals with the possible exception of extremely low temperatures. We shall take into account the following three processes, which will be denoted by superscripts (1,) (2) or (3): (1) Cutting of edge dislocations through the forest of dislocations threading the glide plane, (2) cutting of screw dislocations through this forest, (3) thermally activated generation of vacancies by jogs in screw dislocations. They have this in common, that the activation energy of the process depends linearly on the shear stress τ_1 on the dislocation.* The reason for this simply is that during the activation the dislocation is pushed forward through an ' activation-distance ' d' and the work done during activation is proportional to $d'\tau_1$. In this context τ_1 is to be understood as the shear stress effective on the dislocation line. It is given by the difference of the applied shear stress τ and the shear stress τ_G due to other obstacles to the dislocation movement such as the elastic stress field of neighbouring dislocations. τ_G depends on temperature only indirectly through the shear modulus G.

If only thermally activated processes of one of the two types just mentioned occur the strain rate is given by

$$\dot{\epsilon}=NAb\nu_0 \exp\left\{-\frac{\mathbf{G}_0-v\ (\tau-\tau_G)}{kT}\right\} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

(Seeger 1954 a, 1954 d). N is the density of sites at which the activation takes place, A is the area swept out by the dislocation during and after the activation process, ν_0 is a frequency of the order of magnitude of Debye's frequency, $\mathbf{G}_0$ is the free energy of activation at zero stress, and v is the ' activation volume '. For a dislocation of strength b the activation volume is given by

$$v=bd'l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

l is the distance between the sites at which the dislocation lines are being held up. The activation distance d' has already been introduced. If an isolated vacancy (or an interstitial) is created d' is about equal to the interatomic distance b. The same is true if a dislocation cuts through an unextended dislocation. If the intersected dislocation is an extended one d' will depend on the characters and the orientations of the dislocations involved. d' may be larger or smaller than the separation $2\eta_0$ of the half-dislocations in an extended dislocation (Seeger 1955 b). In Cu, Ag, Au and Ni, $2\eta_0$ is about 10 b, and it seems reasonable to expect also $d'\sim 10\ b$ in these metals.

* For process (3) this holds only if the probability of jumps in backward direction discussed in connection with eqn. (9) is negligible. We therefore expect deviations from the law of eqn. (17) near to and above $T_0^{(3)}$.

Equation (15) neglects the work done by the shear stress acting on the intersected dislocation during the activation process. As this would bring in at best a factor of 2 in v, and in many cases far less than that, we shall neglect this contribution to the activation energy.

The application of the foregoing discussion to the temperature dependence of the flow stress of Al, Mg, Zn and Cd, which are all believed to have high stacking fault energies and therefore only a small separation of the partial dislocations, is discussed elsewhere (Seeger 1954 a, Seeger 1955 b). We shall deal here particularly with Cu, Ag, Au, Ni and Pt which are all thought to have a small stacking fault energy and therefore a wide separation of partial dislocations. If eqn. (15) is solved for the shear stress as a function of strain rate and temperature we obtain

$$\tau=\begin{cases}\tau_G+[H_0-T\,(S+k\ln(NAbv_0/\dot{\epsilon}))]/v & T<T_0\\ \tau_G & T>T_0\end{cases} \quad . \quad . \quad . \quad (17)$$

with

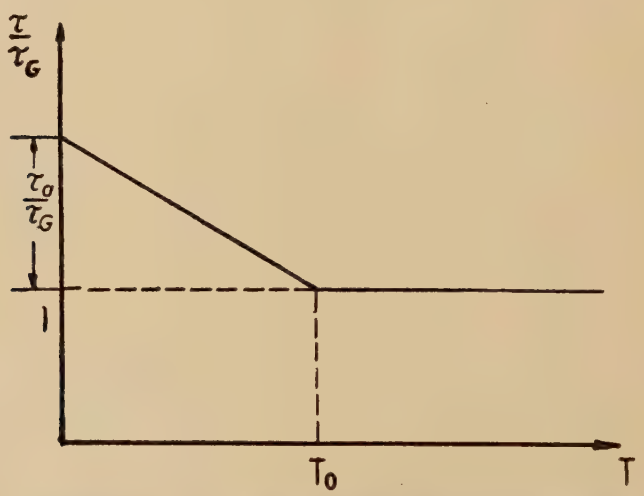
$$T_0=H_0/[S+k\ln(NAbv_0/\dot{\epsilon})]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

In eqns. (11) and (12) we have split the free energy of activation $\mathbf{G}_0$, according to the equation

$$\mathbf{G}_0=H_0-TS, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

into the energy of activation H_0 and the entropy of activation S.

Fig. 8



Schematic representation of the temperature dependence of the flow stress S according to eqn. (17).

The plot of τ (eqn. (17)) versus T is shown in fig. 8. It is characterized by three quantities : the high temperature flow stress τ_G, the temperature T_0, and the flow stress at absolute zero $\tau_G+\tau_a$. Here

$$\tau_a=H_0/v \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

is the ' activation stress ', i.e. the shear stress that has to act on the dislocation line to push it through the obstacle without the aid of thermal energy.

We shall now compare the temperature dependence for the three thermally activated processes under discussion. As Cu is the only member of the group mentioned above for which theoretical estimates of all the quantities determining the temperature dependence of τ are available, we shall confine our present discussion to this metal.

τ_G will be the same for all three processes if the dislocated state of the metal is the same. In testing the theory this has to be secured by using a proper experimental arrangement. $l^{(3)}$ is given by the distance l_j of jogs in a screw dislocation. It will be of the same order of magnitude as $l^{(1)}=l^{(2)}$, which are equal to the distance l'_0 of those dislocations in the forest that are effective in forming obstacles for the gliding dislocations. In experiments carried out after large pre-strains (Cottrell 1954, Wyatt 1953) the forest consists mainly of extended dislocations. For Cu there is both experimental and theoretical evidence that the forest in an unstrained crystal is made up mainly of extended dislocations, too (see Seeger 1954 c). We expect therefore $d^{(1)}=d^{(2)}$ to be about ten times as large as $d^{(3)}$, as discussed above.

For the processes (1) and (2) the free energy of activation under zero stress, $\mathbf{G}_0$, will be taken to be equal to H_0. There may be small entropies of activation $S^{(1)}$ or $S^{(2)}$, but it is even hard to assess their signs. H_0 is made up of two contributions, one from the formation of a constriction (and possibly also a jog) in the intersected dislocation and the other from the formation of a constriction in the gliding dislocation. Both contributions depend on whether the dislocation passing through the forest is an edge or a screw dislocation. $H_0^{(1)}$ is just twice the energy of a constriction in an extended edge dislocation, which was calculated for Cu to be 3·9 ev (Schoeck and Seeger 1955).* $H_0^{(2)}$ is the sum of the energy of a jog in an extended edge dislocation (4·1 ev) and the energy of a constriction in a screw dislocation (0·84 ev). We thus obtain

$$\left.\begin{aligned} H_0^{(1)}&=7{\cdot}8 \text{ ev} \\ H_0^{(2)}&=4{\cdot}9 \text{ ev.} \end{aligned}\right\} \qquad \ldots\ldots \quad (21)$$

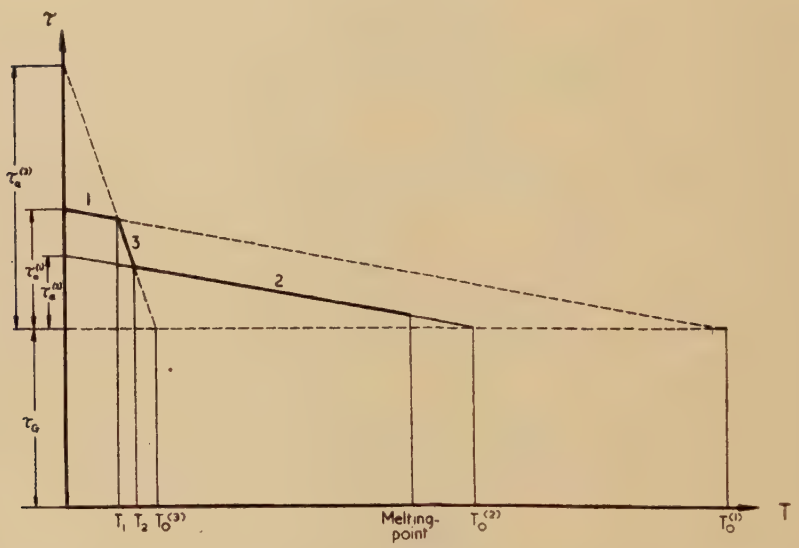
In estimating $\mathbf{G}_0^{(3)}$ we shall take into account both $H_0^{(3)}$ and $S^{(3)}$. In § 2 we have found $H_0^{(3)}=0{\cdot}86$ ev. $S^{(3)}$ takes into account the thermal expansion of the crystal and the increased vibrational entropy of the lattice in the immediate environment of the vacancy. It can be found from the measurements of Meechan and Eggleston (1954) by using the results of Jongenburger (1953) and Abelès (1953) on the electrical resistance of vacancies in Cu. In this way we obtain $S^{(3)}=3{\cdot}4\,k$. (In applying these results to the present problem the assumption is made that $S^{(3)}$ is the same

* The forest is envisaged to consist of both edge and screw dislocations (and also intermediate types which will be neglected). Since the energy of a constriction in an extended edge dislocation is 4–5 times that of an extended screw dislocation (Schoeck and Seeger 1955) edge dislocations form the most effective obstacles in the forest and determine the activation energies. It is, however, the screw dislocations that are responsible for the formation of jogs in the dislocation rings passing through the forest.

whether a vacancy is attached to a dislocation or not. This assumption is probably justified within the accuracy of the present calculation.)

The numerical value of the logarithm in eqn. (17) and eqn. (18) varies somewhat with the experimental conditions under which the flow stress is determined. It depends directly on the strain rate and indirectly (through N and A) on the amount of pre-strain and on the substructure of the crystal. The quantities under the logarithm will be about the same in the three cases considered here with the exception of A, the area swept out by a dislocation line after each activation process. $A^{(1)}$ and $A^{(2)}$ will be the same and larger than $A^{(3)}$ by a factor of the order 10^3–10^4. For slow strain rates, of the order $\dot{\epsilon}=10^{-4}$ sec $^{-1}$, the logarithmic terms will give a

Fig. 9



Temperature dependence of the flow stress of Cu. Only the full parts of the straight lines are observable experimentally.

numerical factor of about 32 for processes (1) and (2) and of about 24 for process (3). These numerical values can either be deduced by choosing plausible values for N, A and ν_0, or by a comparison with experimental results on other metals, notably Zn and Cd (Seeger 1954 d).

Inserting the numerical values the equation for $T_0^{(3)}$ for Cu reads

$$kT_0^{(3)}=\frac{0{\cdot}86}{3{\cdot}4+24}\,\mathrm{ev}=k\times 364^\circ\mathrm{K}. \qquad (22)$$

This temperature is given in the table, together with the corresponding temperatures for the metals treated here. The comparison with experiment will be given below.

For Cu $T_0^{(1)}$ is about 2600°K and $T_0^{(2)}$ is about 1600°K. These temperatures, being higher than the melting point, cannot be observed directly.

The same will be true for the other metals with low stacking fault energies discussed in this paper. In a more indirect way the predictions of this paper can be tested by a comparison with the dependence of flow stress on temperature and strain rate.

At a given temperature the strain rate imposed in a tensile test is maintained by those processes which require the lowest stress for the movement of the required number of dislocations. Figure 9 shows the τ–T relation for the three processes considered in this paper. Since the edge and the screw parts of a dislocation ring move independently of each other the shear stress that will actually be observed is given by the full line in fig. 9 (Seeger 1955 b). As the processes (2) and (3) are prerequisite to each other, the one that requires the higher shear stress will determine the shear stress necessary for the movement of the screw dislocations. If l_j is of the same order as or smaller than l_0' the straight lines corresponding to processes (1) and (3) will intersect and the τ–T relation will show two bends. The reason for this can be seen from eqns. (16) and (20): $H_0^{(1)}/d'^{(1)}$ is slightly smaller than $H_0^{(3)}/d'^{(3)}$, therefore $\tau_a^{(1)}$ will be smaller than $\tau_a^{(3)}$ provided l_j is not larger than l_0'. Figure 9 corresponds to $l_j \approx l_0'/2$.

Adams and Cottrell (1955; see also Cottrell 1954) found experimentally a temperature dependence of the flow stress of pre-strained Cu single crystals of the type shown in fig. 9. The temperatures of the two bends were $T_1 = 180°\text{K}$ and $T_2 = 250°\text{K}$. According to the theory presented here the temperature dependence of the flow stress outside the temperature interval T_1 to T_2 is very small only and comparable with that of the shear modulus. It will therefore be difficult indeed to separate the direct temperature dependence of the flow stress due to the thermally activated processes (1) and (2) from the indirect temperature dependence caused by the temperature variations of τ_G, $H_0^{(1)}$ and $H_0^{(2)}$. These in turn are caused by the temperature dependence of the elastic constants and the stacking fault energy.

It is very much easier to check the prediction of thermally activated processes contributing to the temperature dependence of the flow stress by tests on the change in the strain rate. The experiments of Carreker and Hibbard (1953) on polycrystalline copper give indeed a strain-rate dependence of the flow stress of the right magnitude both below 200°K and above 270°K.

Unfortunately it is not possible to compare the calculated temperatures $T_0^{(3)}$ with experiment, since this temperature is not observable directly. If, however, the temperatures T_1 and T_2 are known from experiment and the ratio $H_0^{(1)}/H_0^{(2)}$ is known from theory, it is possible to find $T_0^{(3)}$.* The evaluation of Cottrell's data gives $T_0^{(3)} = 350°\text{K}$, which compares very well with the theoretical value of 364°K. This close agreement is a strong

* The ratio of $H_0^{(1)}$ and $H_0^{(2)}$ is given to a rather high accuracy by the considerations of §2, although the values for these quantities themselves may be less accurate due to the uncertainty of the exact value of the stacking fault energy. The evaluation of $T_0^{(3)}$ is best done graphically as shown in fig. 9.

indication of the correctness of the theory presented here. Further evidence will be presented in § 4.

§ 4. The Mechanism of Plastic Deformation of Noble Metals

In § 3 we have seen that a straightforward application of the theoretical considerations of § 2 on the generation of vacancies leads to an understanding of the temperature dependence of the flow stress of copper single crystals. We shall now discuss briefly the implications on the mechanism of plastic deformation of noble metals at temperatures below the temperature of self-diffusion. We shall base our arguments on the results on Cu, as there is no reason to suppose that the essential features should differ within the noble metals group. We shall leave out Ni as we do not know at present whether a second bend at the temperature T_1 occurs at low temperatures.

The mechanism below T_1 (in copper $T_1 = 180°\text{K}$) is as described by Seeger (1954 b). The edge dislocations move at lower stresses than the screw dislocations. The form of the dislocation rings spreading across the glide planes shows therefore a marked elongation in the direction of the Burgers vector. Thermal energy is not sufficient to make thermal production of lattice defects an easy process. As the stress is raised during an elongation test the athermal generation of lattice defects as discussed in § 2 will become possible. In elongation tests carried out well below T_1 we therefore expect the defects generated by moving dislocations to be present mainly in the form of interstitial atoms and pairs or clusters of vacancies. There will be a rather small fraction of isolated vacancies only generated at points where l_j is exceptionally large.

Above T_1 the situation is reversed. The thermal production of vacancies, and therefore the spreading of the screw dislocations, is easier than the jog-formation in edge dislocations. The edges will be held up and the shape of the dislocation rings will be elongated perpendicular to the Burgers vector.

In this temperature range many isolated vacancies are created. If the temperature is low enough not to allow appreciable migration of vacancies they are prevented from condensing into clusters or annihilating at dislocation lines. The number of interstitial atoms generated during deformation is of the same order of magnitude as that of vacancies. The reason for this is that jogs creating interstitials are about as frequent as jogs creating vacancies. A screw dislocation stopped by its jogs will move forward preferably at places were vacancies will be created. It will bulge out and generate interstitials by the mechanisms described in § 2.

It can be seen from eqns. (20) and (21) that the stress necessary to move extended edge dislocations through the forest is considerably larger than that necessary to move extended screw dislocations. This cutting of screws through the dislocation forest is the rate determining process in Cu deformed at room temperature. According to Cottrell (1953) the activation energy of such a process can be found from increment tests in logarithmic creep as follows.

In our present notation the equation for the activation energy H_0 at zero stress reads

$$\frac{1}{\tau}\frac{d\tau}{d\ln \nu t} = \frac{kT}{H_0}. \quad \ldots\ldots\ldots \quad (23)$$

Here τ is the applied shear stress which is assumed to be large compared with the stress increment $d\tau$. t is the time and ν a constant with the dimension of a frequency. $d\ln \nu t$ denotes the shift in the t-scale of the creep curve $\epsilon = \epsilon(t)$ to which the raising of the applied stress by $d\tau$ is equivalent. From a series of measurements on polycrystalline Cu at room temperature by Wyatt (1953—in particular table 4) it follows from eqn. (23) that $H_0^{(2)} = 4{\cdot}4$ ev with an accuracy of about 10% in reasonably good agreement with the theoretical value $H_0^{(2)} = 4{\cdot}8$ ev (eqn. 21).* It seems as if the stacking fault energy $\gamma_{Cu} = 40$ erg/cm^2, on which the latter value is based (Seeger and Schoeck 1953) were too low by about 10%.

Cottrell (1954) has observed that for larger pre-strains the only effect on the τ–T curve of changes in the pre-strain is a variation of the τ-scale. The shape of this curve is determined by the magnitude of the lengths l_j, l_0' and l_0–l_0 is a mean distance of the dislocations in all glide systems and determines τ_G (Seeger 1954 b). For larger pre-strains, where slip in more than one glide system occurs, l_0 will be proportional to l_0', which measures the distance of dislocations threading the glide planes. As pointed out earlier, for constant rate of strain l_j will be proportional to l_0'. At larger pre-strains all three lengths involved are proportional to each other and decrease with increasing strain. This is believed to be the explanation of Cottrell's observation.

Another way of comparing the predictions of the theory with experiments is by investigating the magnitude and the annealing behaviour of the additional electrical resistance due to deformation (for surveys see Broom 1954 and van Bueren 1955). The increase of electrical resistance anneals out in four stages (see Appendix B), stages II, III, IV corresponding to the disappearance of point defects created during cold work. Stage V constitutes the additional resistance that does not anneal out below the temperature of recrystallization (in Cu) and is attributed to the scattering of electrons by dislocations.

It is remarkable that after deformation stage IV, corresponding to the migration of vacancies, *is* observed. This shows that Mott's creep mechanism (1955) cannot be the only one capable of creating isolated vacancies, but that there must be at least one more mechanism. This is so because Mott's mechanism requires a temperature which allows the vacancies to anneal out as soon as they are created. As far as the author is aware the mechanism employed in this paper is the only one so far proposed for the generation of single vacancies at low temperatures.

Whereas stage IV is quite pronounced after deformation at room temperature it is, compared with stages II and III, weak after deformation at low

* The strain-rate in Wyatt's increment tests was of the same order of magnitude as that usually applied in tensile test. The comparison of creep data and tensile test appears to be justified.

temperatures and increases somewhat with temperature (preliminary experiments by van Bueren and Jongenburger—private communication). This is in qualitative agreement with the present theory which predicts the generation of a small number of single vacancies at low temperatures. Another effect borne out by these experiments is that stage V, corresponding to the resistance of the dislocation lines, increases with increasing temperature. This is at first sight rather surprising, as the coefficient of work-hardening is practically independent of temperature in the investigated range of temperature and deformation (Blewitt, Coltman and Redman 1955), so no change in the number of dislocations is expected. The experimental results find a natural explanation, however, in terms of the changes occurring in the shape of the dislocation rings with increasing temperature of deformation. At very low temperatures the dislocation rings, being oblonged in the direction of the Burgers vector, consist mainly of screw dislocations, and these are known to have a smaller electrical resistance than edge dislocations (Hunter and Nabarro 1953). The form of the dislocation rings changes with increasing temperature, consisting at room temperature mainly of edge dislocations. The electrical resistance due to the dislocations created during cold work increases therefore with temperature until recovery effects reduce the total number of dislocations appreciably.

It appears from the preceding discussions that for Cu the theory agrees not only with the known facts about mechanical properties but also with the observations on the electrical resistance and its annealing behaviour. In the latter case, however, more complete data are required for a quantitative comparison between theory and observations.

Acknowledgments

The first version of this paper was written during the author's stay in the H. H. Wills Physical Laboratory, University of Bristol, during summer 1954. The author's thanks are due to Professor N. F. Mott, F.R.S. for his hospitality and for valuable discussions. The author would also like to thank Professor Dr. U. Dehlinger for his interest in this work, Dipl. Phys. J. Diehl for illuminating discussions, and Dipl. Phys. H. Stehle for his assistance in analysing the problem of electrical interaction of vacancies and dislocations. The author is grateful to Dr. A. H. Cottrell for comments and for correspondence on his experimental results and to Dr. H. G. van Bueren for communicating the results of his experiments prior to publication and for calling the author's attention to the work of Lazarew and Ovcharenko.

APPENDIX A

Evidence for the Stacking Fault Energy of Nickel and other Transition Metals

As pointed out elsewhere (Seeger 1955 b) the stacking fault energy γ of a f.c.c. or a h.c.p. metal is closely related to its electronic structure. Qualitative conclusions on the magnitude of the stacking fault energy may be

drawn from a study of the factors governing cohesion of these metals. It is, however, hard to obtain quantitative results on a purely theoretical basis as the stacking fault energy per atom does not exceed a few per cent of the binding energy of the metal and is often of the order of only 10^{-3} of the binding energy. It has been possible to group the close-packed metals into metals with high stacking fault energies and metals with low stacking fault energies (for details see Seeger 1955 b). The situation is straightforward with regard to metals without partly filled d-shells. Multivalent metals with f.c.c. or h.c.p. structure belong to the first group (Pb possibly being an exception) ; monovalent metals, in particular the noble metals, belong to the second group.

In the case of transition metals the situation is more complex as both d- and s-electrons have to be taken into account. The author (Seeger 1954 e) has recently undertaken a systematic study of the energetically favourable structures of the transition metals (' energetically favourable ' means in this context that the cohesive energy of the structure differs by not more than, say, 3×10^{-3} from the binding energy of the stable structure). Which of the simple metallic structures are energetically favourable depends on the number of d-electrons per atom. There are some cases in which neither the f.c.c. nor the h.c.p. structure is energetically favourable (V, Nb, Ta, Cr, Mo, W), some others in which the h.c.p. but not the f.c.c. structure is energetically favourable (Ti, Zr, Hf) and yet others in which both the f.c.c. and the h.c.p. structures are energetically favourable (e.g. Fe, Co, Ni, Ru, Rh, Pd, Os, Ir, Pt). If both the close-packed structures are energetically favourable, the stacking fault energy is expected to be low ; if only one of them is favourable, the stacking fault energy will be high.

The known facts on allotropy and stacking fault energies of transition metals are in agreement with these considerations. A further test would possibly be provided by a knowledge of the main glide plane of the hexagonal metals Ru and Os. These metals have subnormal axial ratios (Ru : $c/a=1{\cdot}5824$; Os : $c/a=1{\cdot}5790$) which are of the same order as that of Ti ($c/a=1{\cdot}5873$). Ti is known not to show the basal plane as the main glide plane. This is not surprising, as in such a subnormal structure the basal plane is less densely packed than the $\{10\bar{1}0\}$ and the $\{10\bar{1}1\}$-planes, and (because of the high stacking fault energy of Ti) there is no preference for dislocations in the basal plane. In Ru and Os, however, complete dislocations lying in the basal plane should be able to dissociate freely into partial dislocations and to lower their elastic energy. This energy gain could perhaps be large enough to put the basal plane back in its role as the main glide plane of the hexagonal metals.

Turning now to Ni this transition metal should have a low stacking fault energy. There is no reason why Ni, being situated in the periodic table between Co and Cu, both of which have definitely low stacking fault energies, should have a high one. It appears, however, from the experimental data so far available that the stacking fault energy γ_{Ni} is somewhat

larger than that of Co ($\gamma_{Co}=10$ erg/cm^2 at room temperature) and of Cu ($\gamma_{Cu}=40$ erg/cm^2) (for evidence see below).

The *experimental* evidence for a low stacking fault energy of Ni is as follows :

(*a*) The stacking fault energy of f.c.c. metals is closely related to the boundary energy of coherent {111} twins and about twice as large as the latter. The smaller the ratio of the twin energy to the average grain boundary energy is the more frequent will be the occurrence of recrystallization twins. The fact that annealing twins are common in Ni and that their frequent occurrence makes it virtually impossible to grow single crystals of Ni by the strain-anneal method (P. Haasen, private communication ; E. Kneller, private communication) is therefore an indication that γ_{Ni} is comparatively low. The same situation with respect to the recrystallization twins holds for Cu.

(*b*) It was recently possible to establish experimentally the correctness of the suggestion by Koehler (1952) (see also Leibfried and Haasen (1954), Seeger (1954 f), Haasen and Leibfried (1954), Seeger (1955 b)) that the lamellae within the slip bands of f.c.c. metals originate from glide in one plane by thermally activated screw dislocations (Diehl, Mader and Seeger 1955). The frequency of thermally activated cross-slip depends on the stacking fault energy through the width of extended screw dislocations and also on temperature and applied stress (Schoeck and Seeger 1954). In Al, which has a high stacking fault energy, slip lamellae occur at room temperature already at stresses as low as 10^{-4} G (G=shear modulus). For both Cu and Ni, however, an applied shear stress of about 10^{-3} G is necessary for the formation of slip bands. As the equivalent temperature of Ni is somewhat higher than that of Cu, this experimental result shows that the width of extended dislocations in Ni must be slightly smaller than that in Cu. It appears at present that a stacking fault energy $\gamma_{Ni}=80$ erg cm^2 might be a reasonable estimate.

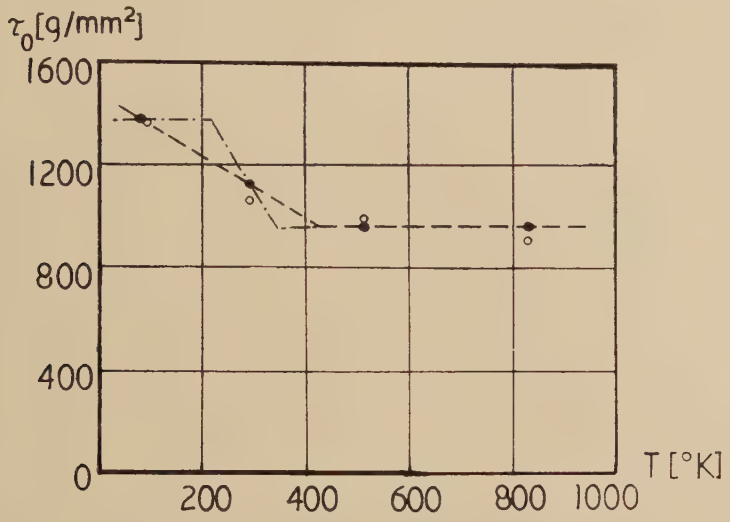
Other experimental data on Ni are on (i) the temperature dependence of the critical shear stress, (ii) activation energy of high temperature steady state creep, and (iii) critical shear stress of the binary alloys Cu–Ni.

(i) The temperature dependence of the critical shear stress has already been discussed in § 3. It was shown there that it fits qualitatively with the present picture of a low stacking fault energy. An interpretation of the data on the basis of a high stacking fault energy (the low temperature part of the dashed curve in fig. 10 being due to jog-formation in edge dislocations) would require that ' extended ' dislocations are as narrow in Ni as they are in Al, and this in turn would require that γ_{Ni} is at least of the order of 400 erg/cm^2. From the point of view of the electronic structure of Ni this seems to be unreasonably large.

(ii) The activation energy of high temperature creep is found to be equal to that of self-diffusion. It is not possible to decide theoretically whether this is due to a very small or to a large splitting of complete dislocations into partials (Seeger 1955 a).

(iii) A recent investigation (unpublished) showed that the analysis of the concentration dependence of the critical shear stress of Cu–Ni alloys by Suzuki (1952) is not quite satisfactory. The resistance to flow due to short range order (Fisher 1954) must not be neglected. Taking into account this contribution to the critical shear stress of the alloys, the twin energy of Ni will come out smaller than given by Suzuki (1952). No final conclusions can be drawn as yet.

Fig. 10



Temperature dependence of the critical shear stress of Ni-single crystals after Andrade and Henderson (1951). o: values given by Andrade and Henderson ; • : extrapolated from the easy glide region according to Seeger (1954 a). The dotted lines show possible interpolations between the measurements.

On the basis of the above evidence it is concluded that Ni has a low stacking fault energy of the order of 80 erg/cm². Since the electronic structure of Pt is very similar to that of Ni we believe Pt, the other transition metal treated in this paper, to have a low stacking fault energy, too.

APPENDIX B

Energies of Lattice Defects in the Noble Metals and in the Transition Metals Ni *and* Pt

In this paper formation and migration energies of lattice defects have been used for the calculation of the temperature T_0[3] and for an identification of lattice defects created during plastic deformation. The values employed for the noble metals have been given elsewhere (Seeger 1955 c). They are based on the classification of annealing stages observed after cold work, radiation damage and quenching as given by van Bueren (1955) and are in agreement with the energies of formation of single vacancies as

calculated by Fumi (1955). Following the Delft school, van Bueren (1955), and others, the annealing stage observed in Cu between 130°K and 200°K (so-called stage II) has been attributed to the migration of pairs and groups of vacancies. Stages III and IV (activation energies in Cu of 0·7 ev and 1·2 ev) are ascribed to the migration of interstitials and vacancies, respectively (following Brinkman, Dixon and Meechan 1954).

The only experimental data on noble metals known to the author which do not fit into this scheme are those of Kauffman and Koehler (1955), which were obtained by quenching and annealing experiments on Au wires. These authors give $(1{\cdot}28\pm0{\cdot}03)$ ev for the formation and $(0{\cdot}68\pm0{\cdot}03)$ ev for the migration of vacancies in Au. The reasons for leaving out these data are as follows: (i) Lazarew and Ovcharenko (1955) performed essentially the same experiments as Kauffman and Koehler and found an energy of formation of vacancies in Au of 0·78 ev. This is in reasonable agreement with the value of 0·67 ev obtained by Meechan and Eggleston (1954), which is used in this paper. The activation energy of migration obtained by Lazarew and Ovcharenko from the annealing kinetics above room temperature is 0·52 ev. This value, which agrees with an earlier value of Kauffman and Koehler (1952) of $(0{\cdot}4\pm0{\cdot}14)$ ev, falls within the activation energies of stage II and must therefore be attributed to the migration of double vacancies. This interpretation is substantiated by the observation that migration and formation energy as found by Lazarew and Ovcharenko do not add up to the activation energy of self-diffusion of 1·96 ev. (ii) The calculations of Fumi (1955) suggest that the energy of formation of a vacancy should be smaller in Au than in Cu. For the latter an activation energy of 0·9 ev appears to be well established. (iii) It appears reasonable to assume that under non-equilibrium conditions a vacancy cannot carry out more than 10^6 jumps until it is either trapped at a dislocation line or combines with another vacancy to form a double vacancy. At temperatures where double vacancies are still very mobile they will anneal out even more rapidly than single vacancies. On this basis it can be shown that the temperatures from which the quenching occurred in the experiments under discussion it was not possible to retain the majority of vacancies.

For the energy of formation of lattice vacancies in nickel we use the value 1·6 ev obtained by Nicholas (1955). In quenching experiments on Pt wires Lazarew and Ovcharenko (1955) found 1·2 ev for the energy of formation of vacancies. From the annealing kinetics after quenching these authors found an activation energy of migration of 1·1 ev. This agrees well with the migration energies of 1·20 ev and 1·19 ev observed by Dugdale (1952) after cold-working and irradiating Pt wires. Unfortunately the activation energy of self-diffusion of Pt does not seem to have been measured. It has been pointed out elsewhere (Seeger 1955 a) that in high-temperature creep by the climbing-mechanism the activation energy Q_c for steady state creep cannot be smaller than the activation energy Q_d of self-diffusion. For metals with low stacking-fault energies (among them Pt—see Appendix A) Q_c should be equal to Q_d. Dorn

(1954) gives for Pt $Q_c = 2{\cdot}4$ ev. It appears therefore that it is correct to attribute to both the formation and the migration of vacancies in Pt activation energies of 1·2 ev.

References

Abelès, F., 1953, *Compt. Rend. hebdom.*, **257,** 796.

Adams, M. A., and Cottrell, A. H., 1955, *Phil. Mag.*, **46,** 1187.

Andrade, E. N. da C., and Henderson, C., 1951, *Phil. Trans.* A, **244,** 177.

Bilby, B. A., 1950, *Proc. Phys. Soc.* A, **63, 191.**

Blewitt, T. H., Coltman, R. F., and Redman, J. K., 1955, *Rep. Conf. Defects in Solids* (London : The Physical Society), p. 369

Boas, W., 1955, *Rep. Conf. Defects in Solids* (London : The Physical Society), p. 212.

Brinkmann, J. A., Dixon, C. E., and Meechan, C. J., 1954, *Acta Met.*, **2,** 38.

Broom, T., 1954, *Advances in Physics*, **3,** 26.

Bueren, H. G., van, 1955, *Z. Metall.*, **46,** 272.

Carreker, R. P., and Hibbard, W. R., 1953, *Acta Met.*, **1,** 654.

Cottrell, A. H., 1952, *L'Etat Solide* (9th Solvay Conference on Physics) (Bruxelles : Stoops), p. 421 ; 1953, *Dislocations and Plastic Flow in Crystals* (Oxford : Clarendon Press), p. 206 ; 1954, *Birmingham Conference on Dislocations.*

Cottrell, A. H., Hunter, C. S., and Nabarro, F. R. N., 1953, *Phil. Mag.*, **44,** 1064.

Diehl, J., Mader, S., and Seeger, A., 1955, *Z. Metall.*, **46,** 650.

Dorn, J. E., 1954, *Symposium on creep and fracture of metals at high temperatures*, Teddington, 1954, paper no. 9.

Dugdale, R. A., 1952, *Phil Mag.*, **43,** 912.

Fisher, J. C., 1954, *Acta Met.*, **2,** 9.

Fumi, F. G., 1954, *Les Électrons dans les Metaux* (10th Solvay Conference on Physics) ; 1955, *Phil. Mag.*, **46,** 1007.

Haasen, P., and Leibfried, G., 1955, *Fortschr. d. Physik*, **2,** 73.

Huntington, H. B., 1953, *Phys. Rev.*, **91,** 1092.

Hunter, C. S., and Nabarro, F. R. N., 1953, *Proc. Roy. Soc.* A, **220,** 542.

Jongenburger, P., 1953, *Appl. Sci. Res.*, **B3,** 237.

Kauffmann, J. W., and Koehler, J. S., 1952, *Phys. Rev.*, **88,** 149; 1955, **97,** 555.

Koehler, J. S., 1952, *Phys. Rev.*, **86,** 52.

Landauer, R., 1951, *Phys. Rev.*, **82,** 520.

Lazarew, B. G., and Ovcharenko, O. N., 1955, *Dokl. Akad. Nauk.*, **100,** 875.

Leibfried, G., and Haasen, P., 1954, *Z. Physik*, **137,** 67.

Meechan, C. J., and Eggleston, R. R., 1954, *Acta Met.*, **2,** 680.

Mott, N. F., 1955, *Nature, Lond.*, **174,** 365.

Nabarro, F. R. N., 1951, *Proc. Roy. Soc.* A, **209,** 278.

Nicholas, J. F., 1955, *Phil. Mag.*, **46,** 87.

Schoeck, G., and Seeger, A., 1955, *Rep. Conf. Defects in Solids* (London : The Physical Society), p. 340.

Seeger, A., 1954 a, *Z. Naturforschg.*, **9a,** 870 ; 1954 b, *Z. Naturforschg.*, **9a,** 758 ; 1954 c, *Z. Naturforschg.*, **9a,** 856 ; 1954 d, *Phil Mag.*, **45,** 771 ; 1954 e, *Les Électrons dans les Metaux* (10th Solvay Conference on Physics) ; 1954 f, 3rd discussion meeting on plastic deformation of metals, Stuttgart ; 1955 a, *Rep. Conf. Defects in Solids* (London : The Physical Society), p. 391 ; 1955 b, *Rep. Conf. Defects in Solids* (London : The Physical Society), p. 328 ; 1955 c, *Naturforschg.*, **10a,** 251.

Seeger, A., and Schoeck, G., 1953, *Acta Met.*, **1,** 519.

Seitz, F., 1952, *Advances in Physics*, **1,** 43.

Suzuki, H., 1952, *Sci. Rep. RITU, Tohoku Univ.*, **4,** 455.

Stroh, A. N., 1954, *Proc. Phys. Soc.* B, **67,** 427.

Wyatt, O. H., 1953, *Proc. Phys. Soc.* B, **66,** 459.

CXXXIII. *Impurity Diffusion in Crystals (mainly Ionic Crystals with the Sodium Chloride Structure)*

By A. B. Lidiard
Atomic Energy Research Establishment, Harwell, Berks.*

[Received July 1, 1955]

Abstract

In this paper elementary lattice kinetic theory is applied to the problem of impurity diffusion by the vacancy mechanism, special attention being paid to the case where there are appreciable attractions between impurity atoms and vacancies. We make use of the approximate methods of association theory, in which we treat impurity–vacancy pairs at small distances of separation ('complexes') on a different footing from pairs at farther separations. Detailed formulae are presented for (i) the diffusion coefficient of impurities in face-centred cubic metals, for which the association model should be very good (Johnson mechanism), and (ii) the diffusion coefficient of divalent ions in polar crystals of the NaCl type. One important aspect of the results on divalent ions in polar crystals is the dependence of the diffusion coefficient $D(c)$ on impurity concentration, c, when there is appreciable association. At low concentrations $D(c)$ is proportional to c but at higher concentrations D increases more and more slowly finally tending to a saturation value. In the absence of appreciable association $D(c)$ remains proportional to c at all concentrations and temperatures. The form of our results shows that experimental determinations of impurity diffusion coefficients are capable of throwing a very direct light on the conditions existing within many of the systems of interest in the field of colour centres (e.g. NaCl and KCl containing alkaline earth chlorides).

§ 1. Introduction

It is now widely appreciated that many of the interesting properties of crystalline solids are only understandable in terms of lattice imperfections or departures from perfect lattice periodicity. (For a general account of this subject see for example Seitz 1952.) Foreign atoms constitute one important type of imperfection and in insulating polar crystals, impurity ions with valency differing from that of the ions of the host lattice are of especial interest for work connected with colour centres (Seitz 1954) and luminescence centres (Runciman 1954). The particular interest in multivalent ions comes from the necessity for compensating the *net* charge that they carry when embedded in the crystal lattice. Thus, to

* Communicated by the Author.

take one of the simplest examples, in NaCl containing substitutional Cd^{2-} ions, the Cd^{2+} ions each carry a net charge $+e$ and this is compensated by the creation of an equal number of Na^{+} ion vacancies each of which has a net charge $-e$ (e is the magnitude of the electronic charge). Thus by the addition of a number of Cd^{2+} ions much greater than the number of intrinsic defects we can obtain a system containing a known constant number of Na^{+} vacancies. This allows properties of the vacancy, such as its mobility, to be studied with far less ambiguity than is possible otherwise. A general complication, however, is the tendency of the Cd^{2+} ions (net charge $+e$) and the Na^{+} vacancies (net charge $-e$) to form stable pairs or ' complexes ' as a result of the attractions existing between opposite charges. Following on earlier work by many others we have previously examined the effects of this association into complexes on the electrical and dielectric properties of these impure salts (Lidiard 1954, 1955 a). The present paper on impurity diffusion represents a continuation of the investigations into the transport properties of these systems. Some of our principal results have already been given in a shorter publication (Lidiard 1955 b). In all these calculations it is assumed that at all the temperatures of interest the impurity remains atomically dispersed and does not coagulate to form colloids. The coagulation of these multivalent impurities has not been much studied so far. Exceptions are provided by the work of Haven (1950) on MgF_2 in LiF, Zückler (1949, quoted by Seitz (1954), p. 21) on $CdCl_2$ in KCl and Miyake and Suzuki (1954) on $CaCl_2$ in NaCl.

For cases such as Cd^{2+} ions in NaCl, where the impurity and the compensating vacancy carry equal but opposite net charges we only need consider aggregates formed from one impurity ion and one vacancy. It is then a good approximation to neglect the interactions of these (neutral) pairs or ' complexes ' amongst themselves or with the unassociated impurity ions and vacancies. In the original theory of association in these crystals as given by Stasiw and Teltow (1947), the mutual interactions of the unassociated impurity ions and vacancies were also ignored. With the help of these approximations the association–dissociation reaction can be described by the mass-action law. Let us adhere to the example of divalent ions in 1 : 1 salts. Let the molar concentration of impurity be c, let the fraction of impurity which has associated be p and let the concentration of thermal defects be negligible ($\ll c$). Then if we only consider association into pairs at the nearest neighbour separation the mass-action law gives us (Lidiard 1954)

$$\frac{p}{(1-p)^2} = cz_0 \exp(\Delta G_0/\boldsymbol{k}T). \qquad . \quad . \quad . \quad . \quad (1.1)$$

Here z_0 is the number of distinct orientations of the complex ; $z_0=12$ in a NaCl type lattice, where either sub-lattice is face-centred cubic, and $z_0=6$ in a CsCl type lattice where either sub-lattice is simple cubic. The quantity ΔG_0 is the Gibbs free energy of association, i.e. the work gained

under conditions of constant pressure and constant temperature in bringing a vacancy from a particular distant position to a particular nearest neighbouring position of an impurity ion. $\boldsymbol{k}$ is Boltzmann's constant.

Now it is possible that the binding energy at the next nearest neighbour and even further separations may also be appreciable and we should then expect the bound vacancy to spend a significant fraction of its time at these remoter separations. We therefore extend the idea of association so as to include complexes in excited states—as we may regard these extended pairs. The calculated thermodynamic properties are rather insensitive to the exact separation at which one draws the distinction between associated and unassociated vacancies (see, for example, a calculation by Fuoss (1934) for the analogous problem in liquid electrolyte solutions). We therefore only need to include excited states of the complex if their energy above the ground state is comparable to or smaller than $\boldsymbol{k}T$. With this extended definition of association, eqn. (1.1) becomes

$$\frac{p}{(1-p)^2} = c\Sigma_i z_i \exp\left(\Delta G_i/\boldsymbol{k}T\right) \quad . \quad . \quad . \quad . \quad (1.2)$$

where z_i is the number of orientations of a complex in the ith excited state with energy $\Delta G_0 - \Delta G_i$ above that of the ground configuration. The factor $\Sigma_i z_i \exp\left(\Delta G_i/\boldsymbol{k}T\right)$ is the partition function for a single complex.

It is assumed that an electric field will alter the relative populations of the various states and orientations of the (dipolar) complexes but that the electrolytic current is contributed solely by the unassociated vacancies. In a previous paper (Lidiard 1955 a) we extended this idea to alternating current phenomena and calculated the response of the complexes and of the free vacancies separately, without regard to the details of the dynamical equilibrium existing between them. This assumption can be supported in the following way. In the definition of association we include excited states until they individually no longer make an appreciable contribution to the partition sum in (1.2). The details of the dynamic equilibrium between the unassociated vacancies and the complexes then only concern one or two of the most highly excited states. Since these states do not contribute appreciably to the properties of the complexes, the effects upon the association–dissociation reaction of those agencies which upset thermodynamic equilibrium (e.g. electric fields, concentration gradients, etc.) will also be negligible. This is the point of view which has been taken in association theory up to now. In § 2 we shall adapt this point of view to a discussion of the problem of impurity diffusion in NaCl type lattices, with the main emphasis on divalent ions diffusing through 1 : 1 salts. At this stage the discussion can be made more general than the Stasiw–Teltow (1947) model, and eqn. (1.2) can be replaced by an equation which allows for the Coulomb interactions among the unassociated ions and vacancies (see, e.g. Lidiard 1954).

This approach may be seen to apply only to systems containing tightly bound complexes, at least one of whose configurations is of such low energy that the mean lifetime of the complex is long compared with the mean times governing the response of the complexes to electric fields, etc. Now in the hypothetical limiting case of only random association the ' associated ' vacancies will contribute to the electrical conductivity as easily as if they were ' unassociated '. Similarly for weakly bound complexes we should expect an electric field to lead to a preferential dissociation, and hence to a contribution to the current proportional to the number of complexes present. The theory which we have discussed so far and which is applied to the diffusion problem in § 2 does not go over into this limit. It needs to be extended by explicit consideration of the association–dissociation reaction in order to include weakly bound complexes. There is a practical interest in this case because the system $NaCl+CaCl_2$ does not show strong association effects (Bean 1952, cf. the system $NaCl+CdCl_2$ as studied by Etzel and Maurer (1950) and by Aschner 1954). In §§ 3, 4 and 5 we shall therefore attempt to discuss the details of the effects which the dissociation–association kinetics have on the observed diffusion constants. Unfortunately we have not yet found it practicable to do so outside the strict framework of the Stasiw–Teltow association model, which assumes that there is an attraction between an impurity ion and a vacancy only when they are nearest neighbours. This is obviously only a very crude approximation to the real state of affairs but it is one which gives a fairly accurate representation of tightly bound complexes in ionic crystals if the unknown parameters are chosen to give the best fit with experiment. It also corresponds exactly with the Johnson (1939) model for the diffusion of substitutional impurities in metals. Accordingly the general equations which we derive in § 3 will first be particularized in § 4 by dropping all terms connected with the net charges on the impurities and vacancies so as to apply to metallic diffusion. In § 5 we apply the general equations to the diffusion of divalent ions in 1 : 1 salts. In all the ionic crystals, we have in mind diffusion on only one of the sub-lattices and we restrict ourselves to NaCl type crystals where the sub-lattices are face-centred cubic. Accordingly our discussion of metallic diffusion applies only to metals with a face-centred cubic lattice.

The principal results for systems containing tightly bound complexes are eqns. (2.11) and (2.12) for the impurity diffusion coefficient. The form of eqn. (2.12) is actually obvious on very general grounds. For. when the complexes have a long lifetime, we can think of them as diffusing particles and the current (in suitable units) is $-D_0\partial(pc)/\partial x$. Referred to the gradient of the *total* impurity concentration this is equal to $-D_0(\partial(pc)/\partial c)(\partial c/\partial x)$ so that the impurity diffusion coefficient has the form (2.12). The main point of § 2 is thus the rigorous derivation of the form for D_0. This has only been done for the NaCl lattice (face-centred cubic sub-lattices) but the extension to other lattices is straightforward.

$D(c)$ for divalent ions in 1 : 1 salts increases proportionally to c for low concentrations, but for larger concentrations $D(c)$ increases more slowly, finally tending to a saturation value, D_0.

The principal results for systems containing complexes of arbitrary lifetime are eqn. (4.4) for face-centred cubic metals, and eqn. (5.6) for divalent ions in 1 : 1 salts of the NaCl type. The shape of the diffusion isotherm (5.6) is roughly the same as (2.12) whatever the lifetime of the complexes. On the other hand as pointed out in § 6, the ionic conductivity isotherms for short-lived complexes will be appreciably different from those appropriate to complexes of long lifetime. The hypothesis that the difference between the observed conductivity isotherms of $NaCl+CdCl_2$ (Etzel and Maurer 1950) and $NaCl+CaCl_2$ (Bean 1952) is largely due to a difference in lifetime (and is not because the Ca^{2+} ions are entirely unassociated) could then be tested experimentally by studying the diffusion of Cd^{2+} and Ca^{2+} ions in these systems.

§ 2. Diffusion of Tightly Bound Complexes

We begin with a discussion of impurity diffusion in systems where the impurity–vacancy complexes are tightly bound but where there are no tightly bound excited states. We consider one-dimensional diffusion parallel to a [100] axis in a NaCl type lattice. The NaCl lattice, being cubic, displays isotropic diffusion properties so that our choice of a [100] direction, although simplifying the calculation, does not restrict the results. Figure 1 shows the diffusion direction and the twelve nearest neighbour positions for a vacancy associated with the impurity atom x. These twelve positions can be divided as shown into three groups (a), (b) and (c) according to the projection of the complex on the diffusion direction (x-axis). The allowed orientations are shown more schematically in fig. 2. We shall refer the orientation of the complex to the position of the impurity ion. Let $an(x)$ be the number of associated impurity ions per unit area of plane x : here a is the shortest anion–cation separation in a NaCl type lattice. Of these associated ions let $an_a(x)$, $an_b(x)$ and $an_c(x)$ have their vacancies in the forward (a), middle (b) and backward (c) directions respectively. Then

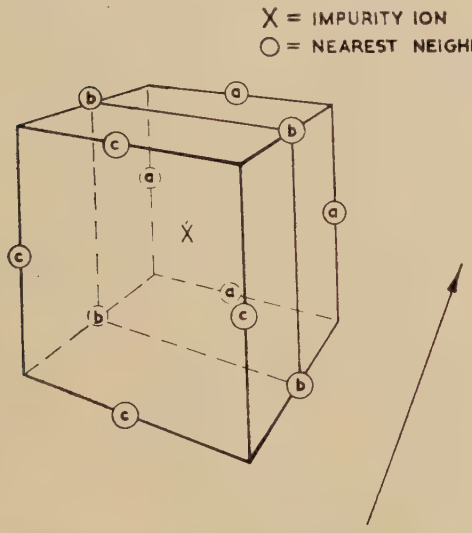
$$n(x)=n_a(x)+n_b(x)+n_c(x).$$

Let w_1 be the probability per unit time that an associated vacancy will jump from one attached position to another particular attached position (one of four possibilities). A complex with a-orientation can jump backwards in this way to either of two b-orientations (and to two other a-orientations) and a complex with c-orientation can jump forwards to either of two b-orientations. Conversely every complex with a b-orientation can jump backwards to one of two c-orientations or forwards to one of two a-orientations. However, direct changes from a to c and c to a are also possible by an exchange of places between the impurity ion and the vacancy. Let this process occur with a probability w_2 per unit time.

On the basis of these definitions we see that the equations for the time rates of change of n_a, n_b and n_c are, *in the absence of any electrical effects*,

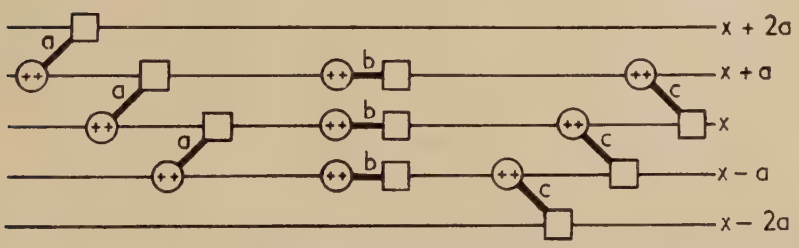
$$\left.\begin{aligned}
\frac{\partial n_a(x)}{\partial t} &= 2w_1 n_b(x) + w_2 n_c(x+a) - w_2 n_a(x) - 2w_1 n_a(x), \\
\frac{\partial n_b(x)}{\partial t} &= 2w_1 n_a(x) + 2w_1 n_c(x) - 2w_1 n_b(x) - 2w_1 n_b(x), \\
\frac{\partial n_c(x)}{\partial t} &= 2w_1 n_b(x) + w_2 n_a(x-a) - w_2 n_c(x) - 2w_1 n_c(x).
\end{aligned}\right\} \quad . \quad . \quad (2.1)$$

Fig. 1



The twelve nearest neighbours to an impurity ion (x) in a face-centred cubic lattice. If the impurity is part of a complex one of these twelve sites will be vacant. When the diffusion direction defines the x-axis as shown then a, b and c represent the planes $x=a$, $x=0$ and $x=-a$. The impurity lies at the centre of $x=0$.

Fig. 2



Schematic diagram showing those orientations and positions of the complexes which must be considered in setting up the equations of §§ 2, 3.

As explained in § 1, when the complexes are tightly bound we can pursue our calculations without regard for the details of the

association–dissociation reaction. This does not necessarily mean that the presence of unassociated vacancies will not be felt in the equations for the motions of the complexes. Thus on the vacancy diffusion model the unassociated impurity ions are unable to move. Their immobility and the mobility of the oppositely charged unassociated vacancies then give rise to a Nernst diffusion potential which prevents the diffusion of the vacancies (see, e.g. Denbigh 1951, especially § 5.3). In fact there arises an electric field just strong enough to oppose the effect of the concentration gradient and reduce the free vacancy current to zero. We must therefore modify eqns. (2.1) by inserting new probabilities appropriate to the case where an electric field $\mathbf{E}$ exists along the diffusion direction. It may be shown that the new probabilities are obtained from w_1 and w_2 by multiplying by an appropriate factor of the form $\exp(q\mathbf{r}\,.\,\mathbf{E}/\boldsymbol{k}T)$, where q is the charge on the ion which jumps (e for parent cations, $2e$ for impurity cations, etc.) and $\mathbf{r}$ is the vector joining the initial equilibrium position of this ion to its position at the saddle point for the transition. (See, for example, Fröhlich 1949, especially Chap. III and Wert and Zener 1949, 1950.) For fields of the magnitude arising experimentally we can generally replace this exponential factor by $1+(q\mathbf{r}\,.\,\mathbf{E}/\boldsymbol{k}T)$. Abbreviating $aeE/\boldsymbol{k}T$ by λ we obtain in place of eqns. (2.1)

$$\left.\begin{aligned}
\frac{\partial n_a(x)}{\partial t} &= 2w_1(1-\tfrac{1}{2}\lambda)n_b(x)+w_2(1-\lambda)n_c(x+a) \\
&\qquad -w_2(1+\lambda)n_a(x)-2w_1(1+\tfrac{1}{2}\lambda)n_a(x). \\
\frac{\partial n_b(x)}{\partial t} &= 2w_1(1+\tfrac{1}{2}\lambda)n_a(x)+2w_1(1-\tfrac{1}{2}\lambda)n_c(x)-4w_1n_b(x). \\
\frac{\partial n_c(x)}{\partial t} &= 2w_1(1+\tfrac{1}{2}\lambda)n_b(x)+w_2(1+\lambda)n_a(x-a) \\
&\qquad -w_2(1-\lambda)n_c(x)-2w_1(1-\tfrac{1}{2}\lambda)n_c(x),
\end{aligned}\right\} \quad . \quad . \quad (2.2)$$

for diffusion in the presence of a field $\mathbf{E}$ along the x-axis. The value of λ, although it will not be needed explicitly, is given by eqn. (5.3) when $k_1=0$ and m_v is the concentration of unassociated vacancies. We shall now suppose that we are only interested in small concentration gradients so that, to the first order of small quantities, eqns. (2.2) become

$$\frac{\partial n_a}{\partial t} = 2w_1(n_b-n_a)+w_2(n_c-n_a)+aw_2\frac{\partial n_c}{\partial x}-2\lambda n_{a0}(w_1+w_2), \quad . \quad (2.3)$$

$$\frac{\partial n_b}{\partial t} = 2w_1(n_a-2n_b+n_c), \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.4)$$

$$\frac{\partial n_c}{\partial t} = 2w_1(n_b-n_c)+w_2(n_a-n_c)-aw_2\frac{\partial n_a}{\partial x}+2\lambda n_{a0}(w_1+w_2). \quad . \quad (2.5)$$

In the terms in λ we have neglected the differences between n_a, n_b and n_c and their (equal) values at equilibrium, n_{a0}. These differences are only

of the order of the concentration gradient, and when multiplied by λ, which is also proportional to the concentration gradient, they give only second order terms—which we have agreed to drop.

An equation for the current density $j_i(x)$ can be obtained by enumerating the numbers of impurity ions entering and leaving plane x and then counting each as contributing half an atom to the current. In this way we find,

$$j_i(x)=\tfrac{1}{2}a\{w_2(1+\lambda)n_a(x)-w_2(1-\lambda)n_c(x+a) \\ -w_2(1-\lambda)n_c(x)+w_2(1+\lambda)n_a(x-a)\}.$$

or to the first order of small quantities,

$$j_i(x)=aw_2(n_a-n_c)-\frac{a^2w_2}{2}\frac{\partial n_a}{\partial x}-\frac{a^2w_2}{2}\frac{\partial n_c}{\partial x}+2\lambda aw_2n_{a0}. \quad . \quad . \quad (2.6)$$

We now wish to derive an equation of the form

$$j_i(x)=-D_0[\partial n(x)/\partial x] \quad \text{(Fick's law)},$$

so that we can deduce an expression for the diffusion coefficient D_0. To do this we notice that the time derivatives $\partial n_a/\partial t$ etc. occurring in eqns. (2.3) to (2.5) are an order of magnitude smaller than any of the other terms appearing in those equations. (The general diffusion equation shows that $\partial n/\partial t$ is, with an appropriate conversion factor, of the same magnitude as $\partial^2 n/\partial x^2$ which we have agreed to neglect by comparison with the other terms in eqns. (2.3) to (2.5).) We shall therefore set the time derivatives $\partial n_a/\partial t$ etc. equal to zero. Equation (2.4) then gives immediately

$$n_b(x)=\tfrac{1}{2}[n_a(x)+n_c(x)]. \quad . \quad . \quad . \quad . \quad (2.7)$$

Substitution into (2.3) and (2.5) then yields,

$$aw_2(\partial n_a/\partial x)=aw_2(\partial n_c/\partial x)=(w_1+w_2)(n_a-n_c+2\lambda n_{a0}). \quad . \quad (2.8)$$

The left-hand equation of this pair requires that

$$n_a-n_c=\text{a constant, } K. \quad . \quad . \quad . \quad . \quad . \quad (2.9)$$

Using (2.6), (2.7), (2.8) and (2.9) it is then a simple matter to show that

$$j_i(x)=-[a^2w_1w_2/3(w_1+w_2)]\partial n/\partial x. \quad . \quad . \quad . \quad (2.10)$$

The *diffusion coefficient of the complexes* is therefore

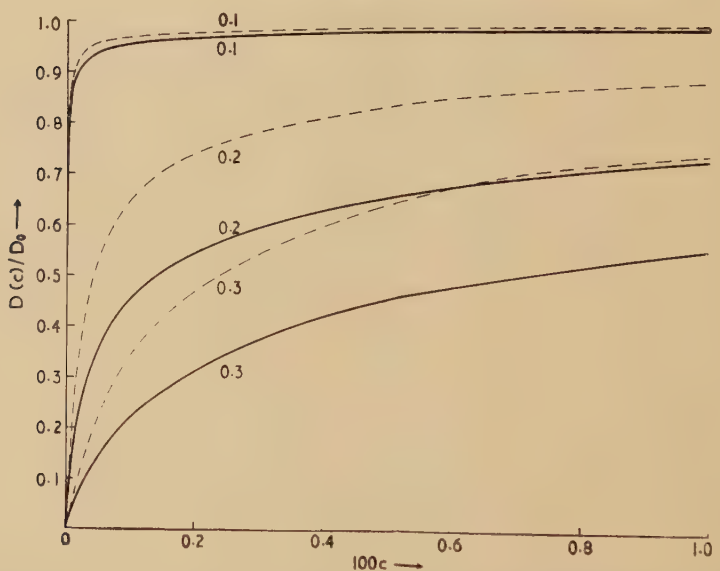
$$D_0=a^2w_1w_2/3(w_1+w_2). \quad . \quad . \quad . \quad . \quad . \quad (2.11)$$

For the continued diffusion of a complex it is necessary that each forward jump of the impurity ion be followed by two forward jumps of the vacancy about the impurity ion as centre. Whichever of these processes is the slower is the one which limits the value of D_0. Equation (2.11) shows that the mean time for a translation of the complex is the sum of the mean times associated with the three necessary jumps, i.e. w_2^{-1}, $(2w_1)^{-1}$ and $(2w_1)^{-1}$.

Although we are here mainly interested in the diffusion of multivalent impurities in NaCl type polar lattices, the assumption of a polar lattice has not really been made in deriving (2.11). We did indeed introduce the Nernst diffusion field, but λ was eliminated from the final equations without having been assigned its special value. The above calculation of D_0 therefore applies equally well to the diffusion of substitutional impurities in face-centred cubic metals when the Johnson (1939) mechanism is operative.

The coefficient D_0 which we have calculated is not of course the diffusion constant for the impurity but only of the complexes. Thus the current in (2.10) is the total impurity current, but the concentration gradient refers only to the complexes. When a local equilibrium is maintained between the associated and the unassociated impurity atoms, we can

Fig. 3



The diffusion coefficient of divalent impurity ions in NaCl type crystals plotted as a function of the fractional molar impurity concentration, for three different temperatures. The full curves are drawn for a model in which the interactions are coulombic at all distances, as treated in an earlier paper (Lidiard 1954). The broken curves are for the same model as calculated with the Stasiw–Teltow approximation. The numbers on the curves represent absolute temperature divided by T_0, where kT_0 is the binding energy of a nearest neighbour complex.

obtain the impurity diffusion constant, $D(c)$—such as we would measure experimentally—by multiplying D_0 by the rate of change of n with total impurity concentration. We get

$$D(c)=D_0[d(pc)/dc] \quad . \quad . \quad . \quad . \quad . \quad (2.12)$$

where p is the degree of association. For impurities in metals p is independent of c (see § 4) and hence D is also independent of c. For divalent ions in ionic crystals the dependence of D on c can be obtained by using the Stasiw–Teltow theory (eqn. (1.1)). Details of the predicted dependence are given in Lidiard (1955 b). However, it is also possible to calculate $D(c)$ by using the more involved equation, connecting p with c, which follows when we allow for the Coulomb interactions existing among the unassociated ions and vacancies. Figure 3 shows diffusion isotherms for a simplified model in which the interactions are coulombic at all distances (*a*) as treated by Stasiw and Teltow, and (*b*) as treated previously by us (Lidiard 1954). We observe that there are no qualitative differences between the two cases.

Before concluding this section we note a further possible extension by inclusion of excited configurations of the complexes. The appropriate model for this case is one where the energies of the excited configurations are only slightly higher than those of the ground configuration ($\Delta G_0 - \Delta G_i \lesssim \boldsymbol{k}T$), but where the association energy of the ground configuration ΔG_0 is much greater than $\boldsymbol{k}T$. This case represents a fairly obvious extension of the simpler equations which we have derived above and will therefore not be considered explicitly here.

§ 3. General Equations for Stasiw–Teltow Association Model

We shall now proceed with our calculations without the restriction to tightly bound complexes, i.e. we shall no longer assume that w_1 and w_2 are large compared with the probabilities for those vacancy jumps which lead to dissociation of the complex. The kinetic equations now become much more complicated, for we must take explicit account of the effects of the concentration gradient and the associated Nernst diffusion field on the association–dissociation reaction. As explained in the introduction these complications have so far restricted our considerations to a crude model in which all pairs at greater separations than the nearest neighbour distance are considered to be non-interacting. Only impurity–vacancy pairs at the nearest neighbour separation are considered to be associated.

Firstly let us consider the geometry of the dissociation jumps which can be made by an associated vacancy. (We are considering motion on a face-centred cubic (sub-)lattice only.) If, starting from the position of the vacancy, we consider in turn all the twelve jump vectors of the type [110] we get twelve new possible positions for the vacancy. One of these corresponds to changing places with the impurity and four correspond to jumps into new nearest neighbour positions. This leaves seven jumps leading to dissociation. If the vacancy was originally in an *a*-position (figs. 1 and 2) we find that four of these seven dissociation jumps have increased the vacancy x-coordinate by a, two have left it unaltered and one has decreased it by a. If the vacancy was originally in a *b*-position the result is different ; two jumps will increase x by a,

three leave it unaltered and two decrease it by a. The case of a vacancy in a c-position is analogous to case a ; four jumps *de*crease x by a, two leave it unaltered and one *in*creases x by a.

Although these final dissociated positions are not all at the same distance from the impurity ion, we shall make the simplifying assumption that jumps into any one of them from an associated position are equally probable ; we let the field free jump probability into a particular dissociated position be k_1 per unit time. We let the probability for the inverse jump be k_2. Jumps of the vacancy which do not involve movement to or from an associated position will be supposed governed by a probability w_0. In the presence of an electric field all the jump probabilities are altered by appropriate factors of the form

$$\exp(q\mathbf{r}\,.\,\mathbf{E}/\boldsymbol{k}T) \doteqdot 1+(q\mathbf{r}\,.\,\mathbf{E}/\boldsymbol{k}T)$$

(see § 2).

We are now in a position to write down the kinetic equations for this system. We let m_i and m_v be the numbers of unassociated impurity ions and unassociated vacancies per unit volume respectively. In terms of the abbreviation λ for $aeE/\boldsymbol{k}T$ we have

$$\begin{aligned}\partial n_a(x)/\partial t = {} & 2w_1(1-\tfrac{1}{2}\lambda)n_b(x)+w_2(1-\lambda)n_c(x+a)-w_2(1+\lambda)n_a(x)\\ & -2w_1(1+\tfrac{1}{2}\lambda)n_a(x)-4k_1(1-\tfrac{1}{2}\lambda)n_a(x)-2k_1n_a(x)\\ & -k_1(1+\tfrac{1}{2}\lambda)n_a(x)+4m_i(x)\{4k_2(1+\tfrac{1}{2}\lambda)[m_v(x+2a)]/N_+\\ & +[2k_2m_v(x+a)]/N_++k_2(1-\tfrac{1}{2}\lambda)[m_v(x)]/N_+\},\end{aligned} \quad . \; . \; (3.1)$$

$$\begin{aligned}\partial n_b(x)/\partial t = {} & 2w_1(1+\tfrac{1}{2}\lambda)n_a(x)+2w_1(1-\tfrac{1}{2}\lambda)n_c(x)-4w_1n_b(x)-2k_1(1-\tfrac{1}{2}\lambda)n_b(x)\\ & -3k_1n_b(x)-2k_1(1+\tfrac{1}{2}\lambda)n_b(x)+4m_i(x)\{2k_2(1+\tfrac{1}{2}\lambda)[m_v(x+a)]/N_+\\ & +[3k_2m_v(x)]/N_++2k_2(1-\tfrac{1}{2}\lambda)[m_v(x-a)]/N_+\},\end{aligned} \quad . \; . \; (3.2)$$

$$\begin{aligned}\partial n_c(x)/\partial t = {} & 2w_1(1+\tfrac{1}{2}\lambda)n_b(x)+w_2(1+\lambda)n_a(x-a)-w_2(1-\lambda)n_c(x)\\ & -2w_1(1-\tfrac{1}{2}\lambda)n_c(x)-4k_1(1+\tfrac{1}{2}\lambda)n_c(x)-2k_1n_c(x)-k_1(1-\tfrac{1}{2}\lambda)n_c(x)\\ & +4m_i(x)\{4k_2(1-\tfrac{1}{2}\lambda)[m_v(x-2a)]/N_++[2k_2m_v(x-a)]/N_+\\ & +k_2(1+\tfrac{1}{2}\lambda)[m_v(x)]/N_+\}.\end{aligned} \quad . \; . \; . \; . \; . \; . \; . \; . \; (3.3)$$

In these equations N_+ is the number of cation (or anion) sites per unit volume ; hence $m_v(x)/N_+$ is the probability of finding an unassociated vacancy at a particular site in lattice plane x. The last term in eqn. (3.1) without the factor 4 is the rate of formation of complexes with a particular a-orientation, by recombination of free vacancies with free impurity atoms. The factor 4 is inserted since there are four distinct orientations in group (a). Similarly for groups (b) and (c) (eqns. (3.2) and (3.3)). The equation for the rate of change of free impurity concentration is likewise.

$$\partial m_i(x)/\partial t = 4k_1(1-\tfrac{1}{2}\lambda)n_a(x) + 2k_1 n_a(x) + k_1(1+\tfrac{1}{2}\lambda)n_a(x) + 2k_1(1-\tfrac{1}{2}\lambda)n_b(x)$$
$$+3k_1 n_b(x) + 2k_1(1+\tfrac{1}{2}\lambda)n_b(x) + 4k_1(1+\tfrac{1}{2}\lambda)n_c(x) + 2k_1 n_c(x)$$
$$+k_1(1-\tfrac{1}{2}\lambda)n_c(x) - 4m_i(x)\{4k_2(1+\tfrac{1}{2}\lambda)[m_v(x+2a)]/N_+$$
$$+[2k_2 m_v(x+a)]/N_+ + k_2(1-\tfrac{1}{2}\lambda)[m_v(x)]/N_+$$
$$+2k_2(1+\tfrac{1}{2}\lambda)[m_v(x+a)]/N_+ + [3k_2 m_v(x)]/N_+$$
$$+2k_2(1-\tfrac{1}{2}\lambda)[m_v(x-a)]/N_+ + 4k_2(1-\tfrac{1}{2}\lambda)[m_v(x-2a)]/N_+$$
$$+2k_2[m_v(x-a)]/N_+ + k_2(1+\tfrac{1}{2}\lambda)[m_v(x)]/N_+\} \quad . \quad . \quad . \quad . \quad (3.4)$$

We could write down a similar equation for the rate of change of m_v by the process of vacancy diffusion and by the dissociation and formation of complexes. But such an equation would have no particular value since additional sources and sinks of free vacancies exist in the crystal in the form of dislocations. These ensure that the concentration of vacancies is everywhere such as to satisfy some additional condition which we have not yet included. In metals (where of course $\lambda=0$) this condition is that the density of free vacancies is everywhere equal to the equilibrium value

$$m_v(x) = \exp(-g_v/\boldsymbol{k}T) = \text{a constant}$$

where g_v is the Gibbs free energy of formation (see e.g. Seitz 1950). In polar crystals containing multivalent impurity ions the corresponding condition is the requirement of electrical neutrality everywhere. For divalent ions in 1 : 1 salts this means $m_i(x)=m_v(x)$. The appropriate condition will be inserted later when we consider particular cases.

Before proceeding it may be noted that eqns. (3.1) to (3.4) neglect the existence of slight interference effects in the jumping probabilities. Thus the probability for a jump of a vacancy to a particular position is really modified by the possibility that there may be a second vacancy or complex already in that position. The equations for the complex re-orientations likewise assume that these motions are not hindered by the presence of other species of the mixture. These effects, being by assumption random, will introduce corrections to w_0, w_1, etc. of the order of the molar concentrations of the various species and are therefore negligible in the systems we are considering.

Following the procedure adopted in § 2 we shall now put eqns. (3.1) to (3.4) into differential form by using Taylor's expansion and retaining only quantities of the first order in the concentration gradient. As remarked previously it is consistent with this to put the time derivatives equal to zero (small departures only from steady state conditions). Furthermore in the terms proportional to λ we set n_a, n_b, etc., equal to their equilibrium values (denoted by suffix 0). With these simplifications we eventually get from eqns. (3.1) to (3.4) the new equations

$$0=-n_a(2w_1+w_2+7k_1)+2w_1n_b+w_2n_c+aw_2(\partial n_c/\partial x)$$
$$+28k_2(m_im_v/N_+)+40ak_2(m_i/N_+)(\partial m_v/\partial x)$$
$$+\lambda\{n_{a0}(-2w_1-2w_2+3k_1/2)+6k_2m_{i0}m_{v0}/N_+\}, \quad . \quad . \quad (3.5)$$

$$0=2w_1n_a-n_b(4w_1+7k_1)+2w_1n_c+28k_2(m_im_v/N_+), \quad . \quad . \quad (3.6)$$

$$0=w_2n_a+2w_1n_b-n_c(2w_1+w_2+7k_1)-aw_2(\partial n_a/\partial x)$$
$$+28k_2(m_im_v/N_+)-40ak_2(m_i/N_+)(\partial m_v/\partial x)$$
$$+\lambda\{n_{a0}(2w_1+2w_2-3k_1/2)-6k_2(m_{i0}m_{v0}/N_+)\}, \quad . \quad . \quad . \quad (3.7)$$

$$0=-12k_2(m_im_v/N_+)+k_1(n_a+n_b+n_c). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.8)$$

This last equation shows that the mass action constant (cf. eqn. (1.1)) for the association reaction remains unaltered provided that the departures from thermodynamic equilibrium are small. Addition of (3.5), (3.6) and (3.7) with use of (3.8) gives

$$\partial n_a/\partial x=\partial n_c/\partial x, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.9)$$

i.e.

$$n_a-n_c=\text{a constant, } K. \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

Also, (3.6) taken with (3.8) yields

$$n_b=(n_a+n_c)/2. \quad . \quad . \quad . \quad . \quad . \quad (3.11)$$

Before finally leaving the general equations and proceeding to a discussion of particular cases we must obtain the general expressions for the impurity ion and the vacancy current densities, j_i and j_v respectively. For the discussion of impurity diffusion in the absence of electrostatic effects we only need j_i. But when we are considering say the diffusion of Cd^{2+} ions through NaCl we need both j_i and j_v, since the diffusion potential gradient (i.e. λ) is determined by the condition that electrical neutrality be maintained locally, i.e.

$$j_i-j_v=0.$$

The calculation of j_i is quite straightforward and the result has been given already (§ 2). Using (3.9) and (3.10), eqn. (2.6) becomes

$$j_i=aw_2K-a^2w_2(\partial n_a/\partial x)+2\lambda aw_2n_{a0}. \quad . \quad . \quad (3.12)$$

On the other hand the evaluation of j_v is rather tedious and is therefore given in an Appendix. The result to the first order of small quantities is

$$j_v=Ka(-2w_1-w_2+3k_1)-4w_0a^2(\partial m_v/\partial x)+a^2(\partial n_a/\partial x)(w_2-20k_1)$$
$$+80a^2k_2(m_v/N_+)(\partial m_i/\partial x)-\lambda an_{a0}(4w_1+2w_2+14k_1)-4a\lambda w_0m_v. \quad (3.13)$$

In arriving at eqn. (3.13) we have made use of eqns. (3.9) to (3.11).

§ 4. Diffusion of Substitutional Impurities in Face-Centred Cubic Metals

In this case there is no diffusion potential and λ is zero. Furthermore the concentration of free vacancies is everywhere maintained at its equilibrium value and $\partial m_v/\partial x$ is also zero. Equation (3.5) with (3.8), (3.10) and (3.11) then gives

$$\frac{\partial n_a}{\partial x}=\frac{\partial n_b}{\partial x}=\frac{\partial n_c}{\partial x}=\frac{K(w_1+w_2+7k_1/2)}{aw_2}. \quad . \quad . \quad (4.1)$$

Equation (3.12) for the impurity current then becomes

$$j_i=-aK(w_1+7k_1/2). \quad . \quad . \quad . \quad . \quad (4.2)$$

To find the diffusion coefficient it only remains to find the gradient of impurity concentration in terms of K. This can be accomplished by combining (3.8) and (4.1). Finally, elimination of K from (4.2) and comparison with Fick's law gives for the impurity diffusion constant

$$D_i=\frac{a^2}{3}\frac{w_2(w_1+7k_1/2)}{(w_1+w_2+7k_1/2)}p \quad . \quad . \quad . \quad . \quad (4.3)$$

where $p=(1+k_1N_+/12k_2m_v)^{-1}$, is the degree of association in this model, i.e. the ratio of the number of impurity complexes to the total number of impurity ions. The degree of association and hence D_i are both independent of impurity concentration; p does however depend on temperature. If, as seems most likely, the free energy of association $\Delta G=\boldsymbol{k}T\ln(k_2/k_1)$ is less than g_v, the free energy of formation of an isolated vacancy, then as $T\to 0$, $p\to 0$ as $12\exp(\Delta G-g)/\boldsymbol{k}T$. We shall discuss two special cases of (4.3):

(i) *Tightly Bound Complexes*

Here $k_1\ll w_1$, w_2, i.e. the complexes have a lifetime $[(7k_1)^{-1}]$ which is long compared with the mean times associated with the w_1 and w_2 jumps. For this case the diffusion constant becomes

$$D_i=\frac{a^2}{3}\left(\frac{w_1w_2}{w_1+w_2}\right)p,$$

i.e. the result (2.11) essentially.

(ii) *Loosely Bound Complexes*

There are several possibilities here but it is instructive to discuss the limit corresponding to an impurity of the same chemical species, i.e. a radio-isotope as ordinarily used in self-diffusion measurements. w_1, w_2, k_1 and k_2 are all equal to w_0, the free vacancy jumping probability, so that for $m_v/N_+\ll 1$, eqn. (4.3) gives

$$D_T=\frac{9}{11}\left(\frac{4a^2w_0m_v}{N_+}\right), \quad . \quad . \quad . \quad . \quad . \quad (4.4)$$

i.e. nine-elevenths of the self-diffusion coefficient as usually calculated for tracers. However, as pointed out by Bardeen and Herring (1952), the

simple calculation of the tracer diffusion coefficient neglects the fact that successive jumps of a given tracer atom are correlated with one another. Thus if a tracer atom has just jumped into a vacancy *its* next jump is not entirely random since it is possible that it will move back into the position it had just left. The factor m_v/N_+ which expresses the assumption of complete randomness in the tracer's jumps must therefore be reduced somewhat. By a rigorous random-walk calculation one can derive* the factor 0·80 for a face-centred cubic lattice, compared with our factor 0·82. The slight difference arises because the present calculation is also incomplete in its handling of the correlation between jumps. We have dealt with nearest neighbour correlations by our use of a rudimentary pair distribution function (numbers of complexes) but we have not taken account of the statistical correlations as they concern pairs at second and farther neighbour separations. Nevertheless it is gratifying that the present calculation is so accurate when applied to the problem of tracer diffusion since the method is designed for systems where the correlation effects are largely due to an attraction between the impurity and the vacancy.

§ 5. Diffusion of Divalent Impurity Ions in a 1 : 1 Salt with NaCl Structure

In this case electrostatic effects play a large part in determining the nature of the general solution. Thus we shall solve eqns. (3.5) etc. by using the condition of electroneutrality. Since the vacancy and the divalent ion carry equal and opposite net charges this condition requires $m_i=m_v$ everywhere and also $j_i=j_v$. Use of this last equation will enable us to fix the value of λ.

The relevant equations are obtained as follows. Firstly eqns. (3.5) and (3.8) (or alternatively (3.7) and (3.8)) yield

$$K(w_1+w_2+7k_1/2)=a(w_2+5k_1)(\partial n_a/\partial x)+\lambda n_{a0}(-2w_1-2w_2+3k_1), \quad . \quad (5.1)$$

where K is defined by (3.10). Also, using (3.8) to simplify (3.13) for j_v and then setting $j_v=j_i$ gives

$$K(2w_1+2w_2-3k_1)=a(2w_2-10k_1)(\partial n_a/\partial x)-4w_0a(\partial m_v/\partial x) \\ -\lambda\{n_{a0}(4w_1+4w_2+14k_1)+4w_0m_v\}. \quad . \quad (5.2)$$

The third equation which we need is eqn. (3.12) for j_i, namely

$$j_i=aw_2K-a^2w_2(\partial n_a/\partial x)+2a\lambda w_2n_{a0}. \quad . \quad . \quad (3.12)$$

In principle we can now eliminate K and λ from these three equations and express the impurity current in terms of $\partial n_a/\partial x$ and $\partial m_v/\partial x$: both

* Bardeen and Herring (1952) give the value 0·90 for this factor. However, as pointed out by LeClaire (private communication to the author), eqn. (A.2) of their paper is incorrect. This should be replaced by

$$f=1+2\theta_1P_\gamma+2\theta_2P_\gamma{}^2+\ldots.$$

When this correction is made the correct result 0·80 is obtained. This has been realized also by Haven (1955).

of which may be related to the gradient of impurity concentration by (3.8). Elimination of K from (5.1) and (5.2) gives an equation for λ, i.e. for the diffusion potential

$$\lambda=-\Phi\{5ak_1(\partial n_a/\partial x)(2w_1+w_2+2k_1) +aw_0(\partial m_v/\partial x)(2w_1+2w_2+7k_1)\}, \quad . \quad . \quad (5.3)$$

where

$$\Phi=\{20k_1n_{a0}(w_1+w_2+k_1)+w_0m_v(2w_1+2w_2+7k_1)\}^{-1}. \quad . \quad (5.4)$$

We can now also find K and hence the current density j_i,

$$j_i=-\{20k_1n_{a0}(w_1+k_1)+w_0m_v(2w_1-3k_1)\}a^2w_2\Phi(\partial n_a/\partial x) -20k_1n_{a0}w_0a^2w_2(\partial m_v/\partial x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.5)$$

We can now make the following substitutions and get the diffusion constant $D(c)$ as a function of the degree of association p, i.e. implicitly as a function of c. We have

$$n_{a0}=\tfrac{1}{3}N_+pc,$$

$$m_v=N_+c(1-p).$$

$$\frac{\partial n_a}{\partial x}=\frac{1}{3}\frac{\partial}{\partial x}(n_a+n_b+n_c)=\frac{1}{3}\frac{\partial(N_+c)}{\partial x}\frac{d(pc)}{dc},$$

$$\frac{\partial m_v}{\partial x}=\frac{\partial(N_+c)}{\partial x}\frac{d}{dc}[(1-p)c].$$

Also, using the fact that on the Stasiw–Teltow association theory (eqn. (3.8))

$$\frac{d(pc)}{dc}=\frac{2p}{(1+p)},$$

we arrive at the following expression for the diffusion constant

$$D(c)=\frac{a^2w_2}{3}\left(\frac{2p}{1+p}\right)\frac{\chi_p(w_1)}{\chi_p(w_1+w_2)}, \quad . \quad . \quad . \quad (5.6)$$

where

$$\chi_p(w_1)=w_0(2w_1+7k_1)-p\{2w_0w_1-k_1(20w_1/3-7w_0)-20k_1^2/3\}.$$

We may note several features of this result.

(1) In the second part of the expression for $D(c)$ the denominator is the same function of (w_1+w_2) as the numerator is of w_1. Hence if $w_2\ll w_1$ then, irrespective of the value of k_1,

$$D(c)=\frac{a^2w_2}{3}\left(\frac{2p}{1+p}\right)=\frac{a^2w_2}{3}\frac{d(pc)}{dc}, \quad . \quad . \quad . \quad (5.7)$$

as is clear on elementary grounds (Lidiard 1955 b). Result (5.7) shows that the simple form for the diffusion constant appropriate to tightly bound complexes (§ 2) will also hold for complexes with any lifetime provided only that $w_2\ll w_1$. Expressed in words this condition states that the mean time for a direct impurity–vacancy exchange should be long compared to the mean time for jumps of the vacancy about the impurity ion as centre. We might expect this to be the case for impurity ions of large radius.

(2) Under conditions of low association energy or high temperatures or small impurity concentrations, the degree of association will be much less than unity. Equation (5.6) then gives

$$D(c)=\frac{a^2w_2(w_1+7k_1/2)}{3(w_1+w_2+7k_1/2)}\,2p, \qquad . \; . \; . \; . \; . \quad (5.8)$$

which may be compared with (4.4). The additional factor 2 in the present case comes from the fact that the degree of association in ionic crystals is proportional to c at low concentrations whereas in metals it is independent of c; hence $d(pc)/dc$ in metals is equal to p but in ionic crystals it is $2p$.

§ 6. Conclusion

In the first part of the present paper (§ 2) we have considered diffusion in systems where the complexes are tightly bound, i.e. systems for which we could proceed without explicit consideration of the association–dissociation reaction. We only studied the case where the ground state of the complex was of much lower energy than any of the excited states, but we indicated how this limitation could be removed. From previously published work (Etzel and Maurer 1950 and Teltow 1949) it seems likely that the calculations for tightly bound complexes are applicable to systems such as $NaCl+CdCl_2$ and $AgBr+CdBr_2$. If this is the case then as pointed out previously (Lidiard 1955 b) measurements of the diffusion of Cd^{2+} ions in these salts are capable of providing information about the degree of association with a directness which is not possible with ionic conductivity measurements. It should perhaps be pointed out however, that direct measurements of the extent of the association are now also possible by application of nuclear resonance techniques (Cohen and Reif 1955).

While the system $NaCl+CdCl_2$ seems to afford a good example of tightly bound complexes, the ionic conductivity measurements made by Bean (1952) on $NaCl+CaCl_2$ indicate the absence of any appreciable association in this system. Partly for this reason we have therefore made an examination of the effects of short lifetime on the predicted properties of the simplified model described in § 3. In general the predicted diffusion isotherm (eqn. (5.6)) is more complicated than is the case for tightly bound complexes. However in the special case where the jump time for the associated impurity ion (w_2) is long compared with the jump time for motion of the vacancy from one associated position to another (w_1)—but without restriction on the lifetime of the complexes —then the diffusion isotherm again takes on a simple form (5.7), easy to relate to experiment.

In this paper we have not explicitly considered effects connected with external applied fields. However it is not difficult to see from eqns. (3.5), (3.8), (3.12) and (3.13) that the d.c. electrical conductivity is

$$\sigma=\frac{4a^2e^2w_0}{\boldsymbol{k}T}N_+c(1-p)+\frac{40a^2e^2k_1(w_1+w_2+k_1)}{3\boldsymbol{k}T(w_1+w_2+7k_1/2)}N_+cp. \quad . \quad (6.1)$$

Only the first term of (6.1) is given by the Stasiw–Teltow theory (see Lidiard 1954) which essentially applies to systems where the complexes have long lifetimes (small k_1). When studying $NaCl+CaCl_2$, Bean (1952) inferred the absence of association from the linearity of the σ vs. c isotherms. The idea here is that the $(1-p)$ factor introduces a curvature into the isotherms. However, we now see that if k_1 is appreciable the presence of the second term in (6.1) may easily be sufficient to bring this curvature back to zero without the necessity for negligible association. Measurements of the diffusion of Ca^{2+} ions through $NaCl+CaCl_2$ should enable us to determine whether k_1 is indeed large for this system. Negligible association, such as Bean inferred, would give a diffusion constant closely proportional to the concentration of Ca^{2+} ions. On the other hand the effects of association on the diffusion isotherms should be clearly visible whatever the value of k_1 (eqn. (5.6) or possibly (5.7)).

Acknowledgment

The author wishes to thank Dr. W. M. Lomer for checking the calculations presented here and for his comments on the manuscript.

References

Aschner, J. F., 1954, *Thesis*, University of Illinois.
Bardeen, J., and Herring, C., 1952, *Imperfections in Nearly Perfect Crystals* (New York : John Wiley), p. 261.
Bean, C., 1952, *Thesis*, University of Illinois.
Cohen, M., and Reif, F., 1955, *Report of the Conference on Defects in Crystalline Solids held at Bristol in July* 1954, p. 44.
Denbigh, K. G., 1951, *The Thermodynamics of the Steady State* (London : Methuen).
Etzel, H. W., and Maurer, R. J., 1950, *J. Chem. Phys.*, **18**, 1003.
Fröhlich, H., 1949, *Theory of Dielectrics* (Oxford : Clarendon Press).
Fuoss, R. M., 1934, *Trans. Faraday Soc.*, **30**, 967.
Haven, Y., 1950, *Rec. Trav. Chim. Pays-Bas*, **69**, 1505 ; 1955, *Report of the Conference on Defects in Crystalline Solids held at Bristol in July* 1954, p. 261.
Johnson, R. P., 1939, *Phys. Rev.*, **56**, 814.
Lidiard, A. B., 1954, *Phys. Rev.*, **94**, 29 ; 1955 a, *Report of the Conference on Defects in Crystalline Solids held at Bristol in July* 1954, p. 283 ; 1955 b, *Phil. Mag.*, **46**, 815.
Miyake, S., and Suzuki, K., 1954, *Acta Cryst.*, **7**, 514.
Runciman, W. A., 1955, *Brit. J. Appl. Phys.*, **6**, S. 78.
Seitz, F., 1950, *Acta Cryst.*, **3**, 355 ; 1952, *Imperfections in Nearly Perfect Crystals* (New York : John Wiley), p. 3 ; 1954, *Rev. Mod. Phys.*, **26**, 7.
Stasiw, O., and Teltow, J., 1947, *Ann. Phys., Lpz.*, **1**, 261.
Teltow, J., 1949, *Ann. Phys., Lpz.*, **5**, 63, 71.
Wert, C., 1950, *Phys. Rev.*, **79**, 601.
Wert, C., and Zener, C., 1949, *Phys. Rev.*, **76**, 1169.
Zückler, K., 1949, *Thesis*, Göttingen University.

APPENDIX

In this appendix we shall obtain the full expression for j_v, the total vacancy current density. The expression which we obtain may then be simplified to apply to conditions where the departures from equilibrium are small, and will then be found to give the formula (3.13) already employed.

We proceed in the same way as in the evaluation of j_i in § 2. That is, we shall enumerate the numbers of vacancies entering and leaving plane x and count each as contributing half an atom to the current. If the vacancy leaves in the forward direction it contributes $+\frac{1}{2}$ to j_v, while if it leaves in the backward direction it contributes $-\frac{1}{2}$ to j_v. Similarly for vacancies which arrive at plane x. Movements of vacancies within plane x may, of course, be ignored.

In the following we need the numbers of complexes, ions etc., in particular lattice planes. From our definitions these are a times the corresponding quantities n_a, n_b, n_c, m_i etc. There is therefore a factor a in j_v. There is also a factor $\frac{1}{2}$ since each jump in or out of plane x contributes either $+\frac{1}{2}$ or $-\frac{1}{2}$ to the current. We therefore remove the factor $\frac{1}{2}a$ at the beginning and assert that j_v is $\frac{1}{2}a$ multiplied by the sum of the terms which we now enumerate.

Firstly consider the terms which come from the re-orientation jumps of the complexes (w_1 and w_2 jumps). From fig. 2 it is clear that they contribute to $2j_v/a$ an amount

$$\begin{aligned}
&-w_2(1+\lambda)n_a(x)-2w_1(1+\tfrac{1}{2}\lambda)n_a(x)\\
&+2w_1(1-\tfrac{1}{2}\lambda)n_b(x)-2w_1(1+\tfrac{1}{2}\lambda)n_b(x)\\
&+w_2(1-\lambda)n_c(x)+2w_1(1-\tfrac{1}{2}\lambda)n_c(x)\\
&-w_2(1+\lambda)n_a(x-a)-2w_1(1+\tfrac{1}{2}\lambda)n_a(x-a)\\
&-2w_1(1+\tfrac{1}{2}\lambda)n_b(x+a)+2w_1(1-\tfrac{1}{2}\lambda)n_b(x-a)\\
&+w_2(1-\lambda)n_c(x+a)+2w_1(1-\tfrac{1}{2}\lambda)n_c(x+a).
\end{aligned}$$

Next consider the terms coming from dissociation of complexes (k_1). We recall that of the seven ways in which an a-complex can dissociate, four lead to an increase of a in the vacancy's x-coordinate, two leave it unaltered and one decreases x by a. For a b-complex two of the dissociation jumps increase the vacancy x by a, three leave it unaltered and two decrease it by a. Lastly for a c-complex only one dissociation jump increases the vacancy x-coordinate by a whereas two leave it unaltered and four decrease it by a. We then see that dissociation jumps contribute to $2j_v/a$ an amount

$$\begin{aligned}
&-k_1(1+\tfrac{1}{2}\lambda)n_a(x)+4k_1(1-\tfrac{1}{2}\lambda)n_a(x-a)\\
&-k_1(1+\tfrac{1}{2}\lambda)n_a(x-a)+4k_1(1-\tfrac{1}{2}\lambda)n_a(x-2a)\\
&-2k_1(1+\tfrac{1}{2}\lambda)n_b(x+a)+2k_1(1-\tfrac{1}{2}\lambda)n_b(x)\\
&-2k_1(1+\tfrac{1}{2}\lambda)n_b(x)+2k_1(1-\tfrac{1}{2}\lambda)n_b(x-a)\\
&-4k_1(1+\tfrac{1}{2}\lambda)n_c(x+2a)-4k_1(1+\tfrac{1}{2}\lambda)n_c(x+a)\\
&+k_1(1-\tfrac{1}{2}\lambda)n_c(x+a)+k_1(1-\tfrac{1}{2}\lambda)n_c(x).
\end{aligned}$$

Thirdly we find the contribution to $2j_v/a$ coming from jumps of free vacancies into associated positions. Free vacancies at x can contribute both by jumping forwards and backwards. We get

$$\begin{aligned}
&m_v(x)\times 16k_2(1-\tfrac{1}{2}\lambda)m_i(x+2a)/N_+ && (c \text{ at } x+2a)\\
&m_v(x)\times\ 8k_2(1-\tfrac{1}{2}\lambda)m_i(x+a)/N_+ && (b \text{ at } x+a)\\
&m_v(x)\times\ 4k_2(1-\tfrac{1}{2}\lambda)m_i(x)/N_+ && (a \text{ at } x)\\
-&m_v(x)\times\ 4k_2(1+\tfrac{1}{2}\lambda)m_i(x)/N_+ && (c \text{ at } x)\\
-&m_v(x)\times\ 8k_2(1+\tfrac{1}{2}\lambda)m_i(x-a)/N_+ && (b \text{ at } x-a)\\
-&m_v(x)\times 16k_2(1+\tfrac{1}{2}\lambda)m_i(x-2a)/N_+ && (a \text{ at } x-2a).
\end{aligned}$$

Free vacancies at $(x-a)$ can only contribute through jumping forwards, giving

$$\begin{aligned}
&m_v(x-a)\times 16k_2(1-\tfrac{1}{2}\lambda)m_i(x+a)/N_+ && (c \text{ at } x+a)\\
&m_v(x-a)\times\ 8k_2(1-\tfrac{1}{2}\lambda)m_i(x)/N_+ && (b \text{ at } x)\\
&m_v(x-a)\times\ 4k_2(1-\tfrac{1}{2}\lambda)m_i(x-a)/N_+ && (a \text{ at } x-a).
\end{aligned}$$

In the same way free vacancies at $(x+a)$ can only contribute by jumping backwards. They give contributions to $2j_v/a$ of

$$\begin{aligned}
-&m_v(x+a)\times 16k_2(1+\tfrac{1}{2}\lambda)m_i(x-a)/N_+ && (a \text{ at } x-a)\\
-&m_v(x+a)\times\ 8k_2(1+\tfrac{1}{2}\lambda)m_i(x)/N_+ && (b \text{ at } x)\\
-&m_v(x+a)\times\ 4k_2(1+\tfrac{1}{2}\lambda)m_i(x+a)/N_+ && (c \text{ at } x+a).
\end{aligned}$$

The last terms to be enumerated are those which come from the motion of free vacancies from one unassociated position to another. They are

$$\begin{aligned}
&4w_0(1-\tfrac{1}{2}\lambda)m_v(x)-4w_0(1+\tfrac{1}{2}\lambda)m_v(x)\\
&4w_0(1-\tfrac{1}{2}\lambda)m_v(x-a)-4w_0(1+\tfrac{1}{2}\lambda)m_v(x+a).
\end{aligned}$$

The sum of all these terms which we have written down is $2j_v/a$. In asserting this we are omitting any interference terms. Thus the number of free vacancies jumping to new unassociated positions should strictly take account of the fact that not all final positions are equally likely, owing to the presence of the other ions and vacancies. However, this fact will introduce a correction to w_0 which is only of the order of the impurity concentration and such corrections may therefore be neglected. To the first order of small quantities, j_v may then be shown to reduce to eqn. (3.13).

CXXXIV. *The 27-day Recurrence Tendency of Cosmic Ray Intensity*

By I. J. van Heerden* and T. Thambyahpillai †
University of Manchester ‡

[Received August 23, 1955]

Abstract

The 27-day recurrence tendency of the cosmic ray intensity has been studied for the nucleon, meson and electron components of the cosmic radiation and the amplitude of the variation is found to be greater for the nucleon component, furthermore the changes in cosmic ray intensity seem to precede those of magnetic activity by 4 or 5 days. The 27-day period corresponds to the time of rotation of the sun and this would seem to indicate that the cosmic ray intensity changes are associated with solar disturbances which, after a delay of several days, produce magnetic storms. There is some evidence to show that the recurrence tendency is due to decreases in intensity which occur at 27-day intervals.

§ 1. Introduction

It is now well established that the cosmic ray intensity shows a marked 27-day recurrence tendency, similar to that shown by geomagnetic storms of moderate intensity. The continuous observation of intensity variations using ionization chambers and counter telescopes is limited to primary particles with energies above about 10^{10} ev. It is also possible to study the time variations in a lower energy region. The cosmic ray neutron flux at sea-level is largely produced by latitude sensitive primaries, i.e. by primary particles of energy between 10^{9} and 10^{10} ev. Simpson *et al.* (1953) have shown that the intensity variations in this energy region can be studied by means of neutron monitors.

From June 1952 to June 1954 continuous measurements of the cosmic ray intensity have been made at Manchester using a neutron pile monitor, and two arrays of Geiger–Müller counters inclined at 45° to the vertical in the North and South directions. It has therefore been possible to compare directly the intensity variations found in the nucleon component of the cosmic radiation with those found in the ionizing component. In this paper the experimental results on the 27-day recurrence tendency in cosmic ray intensity are reported.

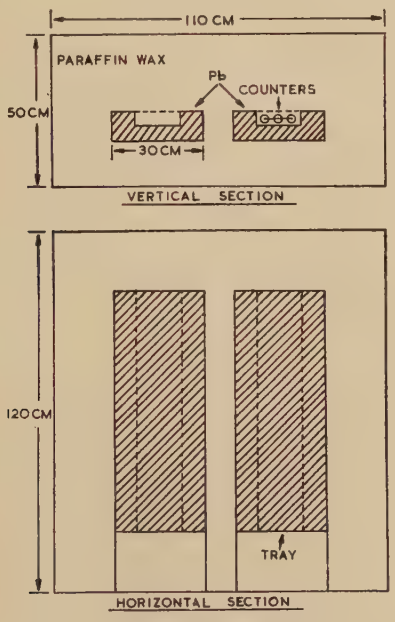
---

* On leave from National Physical Laboratory, Pretoria, South Africa.
† Now at University of Ceylon, Ceylon.
‡Communicated by Professor P. M. S. Blackett, F.R.S.

§ 2. The Neutron Detector

The details and dimensions of the neutron pile monitor are shown in fig. 1. The entire pile was housed in a thermally insulated hut, and the temperature was kept constant by thermostatically controlled heaters. The counters used were of the copper-in-glass cathode type with a central anode wire of tungsten. The effective length of each counter was 50 cm and the diameter 4 cm. They were filled with BF_3 gas enriched with the B–10 isotope at 40 cm Hg pressure, and operated in the proportional region. Three counters were connected in parallel and directly coupled to a high-gain linear amplifier. Output pulses from the amplifier were selected by a discriminator circuit which also provided the correct pulse shape for the scaling circuit. The pulses were scaled by 16 before being fed to a drive unit which operated electro-mechanical registers. These registers, together with a clock, an aneroid barometer and a mains voltmeter were mounted on a panel, which was automatically photographed every 15 minutes.

Fig. 1



Cross section of neutron pile.

The apparatus was designed to contain two trays of counters operating colaterally, but the experimental results described in this paper were obtained with only one set of counters installed. The counting rate was about 2000 per hour.

To check its overall performance, the neutron monitor was calibrated periodically by placing a (Ra, Be) source at a fixed position near the

centre of the pile. By subtracting the number of neutrons generated in the pile by the cosmic radiation from the number recorded with the source in position, the instrumental efficiency for the detection of neutrons could be determined. From June 1952 to March 1953 the efficiency gradually increased by about 4% of its original value. From April 1953 to June 1954 the change has been less rapid.

§ 3. Experimental Results

(*a*) *The Influence of Meteorological Variables on the Neutron Intensity Measured at Sea-level*

If the number of neutrons locally produced in a neutron pile, were related to the primary radiation only by secondary nucleons, the neutron intensity would depend only on the atmospheric pressure, because of the well-known mass absorption effect. The propagation of the nucleon component through the atmosphere is, however, more complicated than this, since secondary mesons can give rise to further high energy nucleon production in the atmosphere, or star production within the pile. The presence of these unstable links could therefore result in a dependence of the intensity of the nucleon component at sea-level on atmospheric temperature as well as pressure.

Table 1

Group	800 mb	500 mb	200 mb	100 mb
1	− 0·08	−0·01	−0·02	−0·10
2	− 0·21	0·0	+0·24	+0·05
3	− 0·26	−0·29	−0·29	−0·23
4	+ 0·07	+0·01	+0·23	−0·10
Mean	−0·12	−0·07	+0·04	−0·10

To determine whether the neutrons had any appreciable temperature effect, a partial correlation analysis was carried out for a total of 118 days during the period June 1952 to October 1952. The variables used were the daily neutron intensity N, the mean atmospheric pressure B, and the mean height H of different pressure levels in the atmosphere. The upper air data were taken from the radio-sonde measurements of the Liverpool Meteorological Station. The partial correlation coefficients $r_{NH,B}$ obtained from the analysis are shown for four monthly groups of data in table 1.

For all the pressure levels considered $r_{NH,B}$ had a very small value, and never reached the 0·1 level of significance. It was therefore assumed that the neutron intensity depends only on the atmospheric pressure, and the daily neutron intensity was correlated with the mean atmospheric pressure for twelve monthly periods from June 1952 to June 1953. The

mean value of the correlation coefficient was −0·97, and the weighted mean of the barometric coefficient was

$$\beta_{NB} = -10{\cdot}4 \pm 0{\cdot}1\% \text{ per cm Hg.}$$

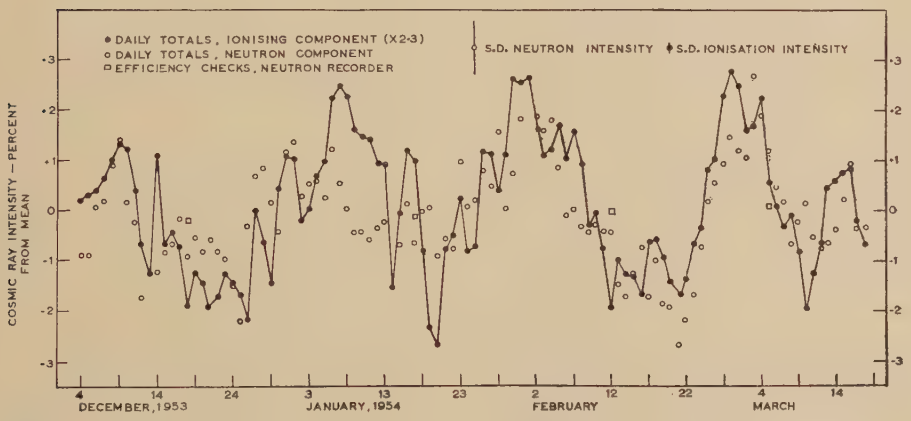
The error given is the standard deviation calculated from the scatter of the individual values about the mean. The absorption mean free path in air for the nucleon component as calculated from the barometric coefficient is

$$L = 130 \pm 1 \text{ g cm}^{-2}.$$

(b) The Relative Amplitudes of the Recurrence Tendency Measured by Different Detectors

Using barometric coefficients of −10·4% per cm Hg for the daily neutron intensity, and −2·64% per cm Hg for the ionizing intensity, the data obtained at Manchester during the period June 1952 to June 1954 were corrected for fluctuations in atmospheric pressure. The daily intensity totals were then plotted as percentage deviations from the mean intensity. As an example of the results obtained, fig. 2 shows the

Fig. 2



Day-to-day changes in cosmic ray intensity from December 4, 1953 to March 18, 1954.

day-to-day changes in the cosmic ray intensity for the period December 4, 1953 to March 18, 1954. Local time midday-to-midday totals were used. The variations in the ionizing intensity, for which the mean of the two directional telescopes were used, were multiplied by a factor of 2·3. The reason for this will be discussed later. The standard deviations of the daily intensity totals due to random fluctuations were ∼0·5% for the neutron and ∼0·1% for the ionizing intensity.

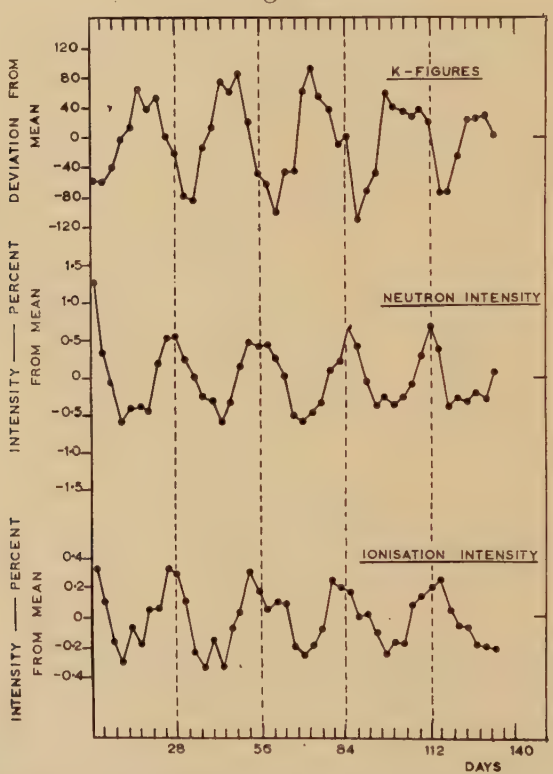
Because of the rather large statistical uncertainty of the daily neutron intensity, it is not possible to be definite about the day on which a maximum or a minimum in the neutron intensity occurred. Nevertheless, it is obvious that :

(1) Although the neutron detector responds to low energy primaries, and the ionization detector to relatively higher energy primaries, there is a fairly good agreement between the intensity changes measured by the two types of detectors. Perfect agreement cannot be expected since no temperature correction has been applied to the data for the ionizing component.

(2) The intensity records for the period under consideration show a remarkable recurrence of maxima and minima with each series having a period of approximately 27–28 days. For example, the maximum occurring on December 9, 1953 is followed by other maxima near January 7, February 1 and February 28, 1954.

(3) Occasionally more than one series of 27-day variations may be present. Thus the smaller maximum occurring near December 27, 1953 is followed by fairly prominent maxima near January 23, February 18 and March 15, 1954.

Fig. 3



Recurrence diagrams for *K*-figures, neutron intensity and ionization intensity for period June 20, 1952 to July 20, 1953.

This recurrence tendency could be seen in all the cosmic ray intensity records, and a rigorous analysis was therefore carried out to determine the period and amplitude. The superposed-epoch method developed by Chree (1913) was used in this analysis. Days of maximum and minimum

intensity were selected at the rate of about one per month. Plotting the average intensity as a function of the day-number, showed a recurrence tendency in the maxima and minima of both the neutron and ionizing intensities. For the purpose of smoothing, the recurrence diagram obtained by using minima as zero days, was subtracted from that obtained by using maxima as zero days.

Fig. 4

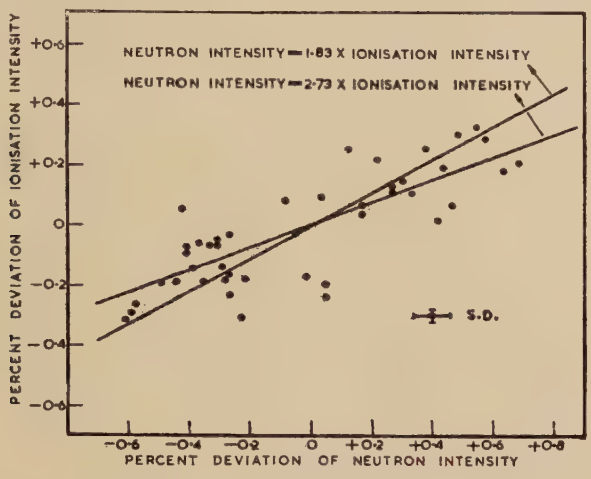
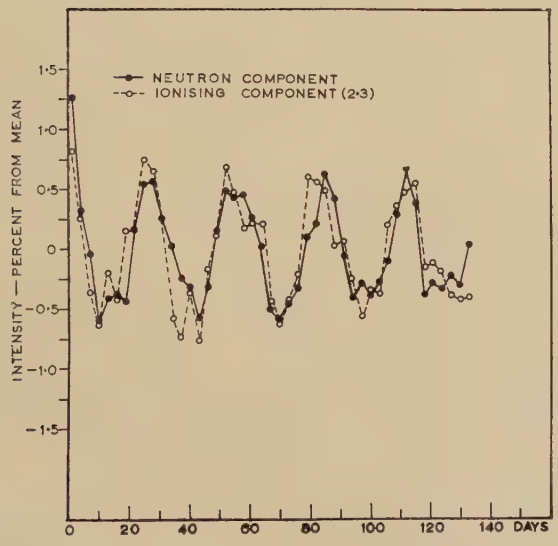


Fig. 5



Recurrence diagrams for neutron and ionizing intensities. The deviations from the mean of the ionizing intensity have been multiplied by 2·3. Data used covered the period June 20, 1952 to July 20, 1953.

The difference curves obtained from a Chree analysis of the daily neutron and ionizing intensities for the period June 20, 1952 to July 20, 1953 are shown plotted in fig. 3. The difference curve obtained from a Chree

analysis of the K-magnetic figures at Abinger is also shown plotted in fig. 3. The same zero days have been used for all these diagrams. It is obvious that both the neutron and ionizing components exhibit a 27–28 day recurrence tendency.

To determine the relative amplitudes of the 27-day recurrence tendency for the neutron and the ionizing components the percentage deviation from the mean of the ionizing component for each day-number was plotted against the corresponding data for the neutrons. These points, for which the mean deviations over three consecutive days were used, are shown plotted in fig. 4. The correlation coefficient between the two sets of data is 0·82. The slopes of the two lines drawn in fig. 4 represent the limits of the regression coefficient. These limits, which result if infinite weight is given to one or the other of the two co-ordinates, are 1·83 and 2·73. Equal weights were assigned to the two co-ordinates, and a value 2·30 obtained for the regression coefficient by the method of least squares. To check this the Chree diagrams, upon which the above analysis was based, have been plotted again in fig. 5 with the percentage deviations from the mean of the ionizing intensity multiplied by a factor of 2·3. There is a very good agreement between the sets of results. The 27-day neutron variations are therefore about twice as large as the variations in the ionizing component.

In addition some earlier data obtained by Elliot and Dolbear (1950) on the hard component, and by Dawton (1952) on the hard and soft components have been analysed using the superposed epoch method. Elliot and Dolbear made their measurements with two directional counter telescopes which were shielded by 35 cm of lead. Dawton's measurements were made by using two similar counter arrays, one containing no absorber and the other containing 10 cm of lead. The unshielded array measured the total flux, the shielded array the intensity of the hard component, and the difference in the rates gave a measure of the soft component intensity. Figure 6 (*a*) shows the Chree diagrams obtained from Elliot and Dolbear's measurements on the hard component. The intensity records obtained between January 5, 1949 and May 3, 1950 were used in this analysis. There is a good similarity in the points obtained for the North and South directional records, the correlation coefficient being 0·82. The approximately 27-day recurrence tendency is quite apparent, and seems to persist on the average for at least four solar rotations, with an amplitude of about 0·2%. The points plotted in fig. 6 (*b*) were obtained from Chree analyses of Dawton's measurements on the hard and soft components. The diagram drawn in this figure is the Chree diagram for the total cosmic ray intensity. The above analyses covered the period May 9, 1950 to April 30, 1951. There are two points worthy of note :

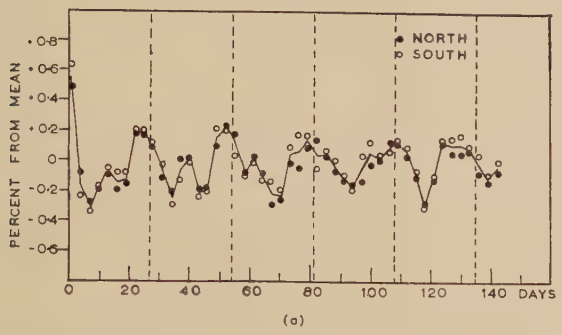
(1) There is a good agreement between the total T, hard H, and soft S intensities, and the correlation coefficients were found to be

$$r_{T,H}=0{\cdot}97, \quad r_{T,S}=0{\cdot}91, \quad r_{H,S}=0{\cdot}85.$$

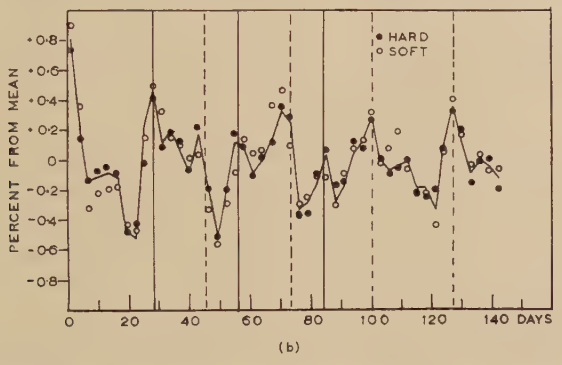
These values all reached the 0·01 level of significance. The limits of the regression coefficients are

$$0{\cdot}98<\beta_{T,H}<1{\cdot}05, \quad 0{\cdot}79<\beta_{T,S}<0{\cdot}95, \quad 0{\cdot}73<\beta_{H,S}<1{\cdot}01.$$

Fig. 6



Recurrence diagrams for cosmic ray intensity in the North and South directions. The measurements have been made using telescopes shielded by 35 cm lead. Data used covered period from January 5, 1949 to May 3, 1950.



Recurrence diagram for the total cosmic ray flux. The points shown plotted were obtained from Chree analyses of the hard and soft component intensities. Data was from May 9, 1950 to April 30, 1951.

With equal weights assigned to the different sets of data, the method of least squares gave the following values for the regression coefficients :

$$\beta_{T,H}\sim 1{\cdot}0, \quad \beta_{T,S}\sim 0{\cdot}87, \quad \beta_{H,S}\sim 0{\cdot}87.$$

In view of the large statistical errors involved in the measurements of the soft component intensity, however, it is doubtful whether any significance can be attached to the departures from unity of the regression coefficients, and it therefore seems very likely that the 27-day recurrence tendency has the same amplitude for the total, hard and soft intensities.

(2) There are two distinct series of peaks present. The first series persists for three solar rotations with peaks of diminishing amplitude at 28,

56 and 84 days. The second series has a small peak at 46 days, followed by three definite peaks at 73, 100 and 127 days (corresponding to the dotted lines).

From the above it is obvious that the neutron intensity as well as the hard and soft component intensities show a 27-day recurrence tendency, and that the ampitude thereof is the same for the total cosmic ray flux, the hard and soft components, and approximately one-half that of the neutron component.

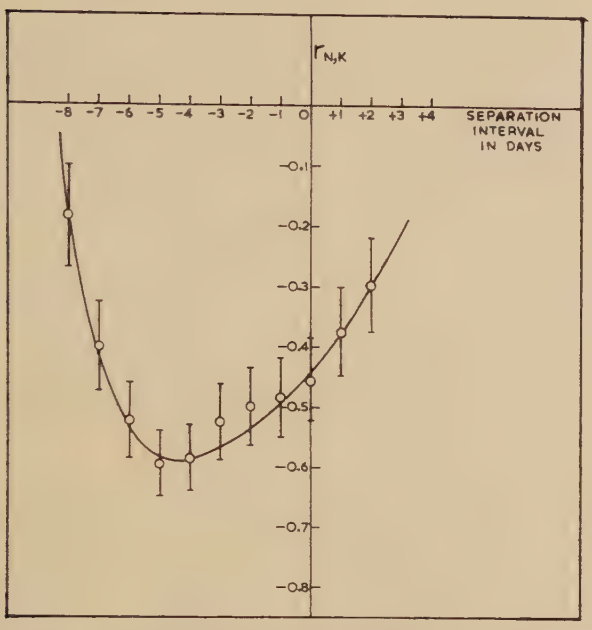
(*c*) *The 27-day Recurrence Tendency and Geomagnetic Field Variations*

Moderate geomagnetic disturbances are known to possess a pronounced 27-day recurrence tendency. This recurrence tendency is clearly shown in the Chree diagram for the K-figures shown in fig. 3. It is also obvious from fig. 3 that there is a strong negative correlation between the Chree diagram for the K-figures, and those for the neutron and ionizing intensities. It is known that decreases in intensity of up to 8% can occur at the time of intense geomagnetic storms and the relationship shown in fig. 3 would seem to be of the same kind, viz. an intensity decrease at the time of a magnetic disturbance. However, there appears to be a difference in detail. Inspection of the K-figure and cosmic ray intensity plots in fig. 3 shows that the two phenomena are not quite inversely related, i.e. the time of minimum cosmic ray intensity is not quite the same as for maximum K-figure.

In order to determine more precisely the relationship between the 27-day recurrence tendency of the cosmic ray intensity and that for the geomagnetic activity, the following procedure was used. The correlation coefficients between the neutron and ionizing intensities and K-figure diagrams shown in fig. 3 were evaluated, as the magnetic data was moved by an integral number of days in the positive or negative direction relative to the cosmic ray data. For instance, for a separation interval of -1 day, the neutron intensity on the zero day, D_0, i.e. N_0, was correlated with the magnetic activity on day D_{+1}, i.e. K_{+1}, N_{+1} with K_{+2} and so on. For a separation interval of $+1$ day the neutron intensity on day D_{+1}, i.e. N_{+1} was correlated with the magnetic activity on day D_0, i.e. K_0, N_{+2} with K_{+1} and so on. Figure 7 shows a plot of the correlation coefficients obtained for different separation intervals in the case of the neutron intensity, whereas fig. 8 shows the corresponding data for the ionizing intensity. Maximum correlation occurs in both cases for a separation interval of about -5 days, indicating that the changes in cosmic ray intensity precede those of magnetic activity by about 5 days.

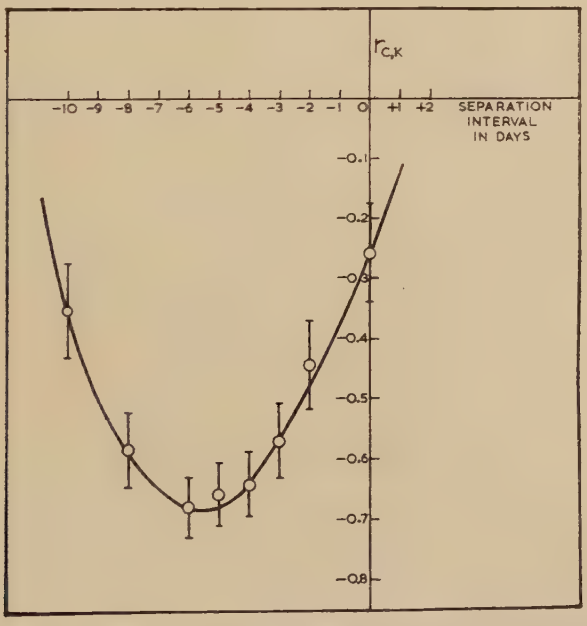
Forbush (1954) has also studied the relation between cosmic ray decreases and various measures of geomagnetic activity, and has shown that these intensity decreases occur during magnetic disturbances. On the other hand, Simpson (1954) and Kane (1955) have put forward some evidence indicating that the maxima of the cosmic ray intensity precede the

Fig. 7



Correlation coefficients between neutron intensity and K-figures for different separation intervals.

Fig. 8



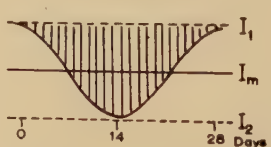
Correlation coefficients between ionizing intensity and K-figures for different separation intervals.

K_p maxima by a few days, and they conclude that the most probable time for the occurrence of a geomagnetic disturbance is near the beginning of a cosmic ray intensity decrease. The high degree of negative correlation between the cosmic ray intensity and the geomagnetic activity in the above results does not necessarily mean that the cosmic ray intensity changes are causally related to the geomagnetic field disturbances. The two effects may have a common cause. Simpson (1954) has shown that the day-to-day variations in the neutron intensity at latitudes above the knee of the intensity vs. latitude curve were very similar to those observed at low latitudes, and he therefore concluded that the 27-day recurrence tendency of the neutron intensity and of the highly correlated ionizing intensity is not produced by geomagnetic field variations. The 27-day recurrence corresponds in time to the rotation of the solar equatorial latitudes. The present results therefore suggest that the cosmic ray intensity changes are associated with solar disturbances, which, after a delay of several days, produce magnetic storms.

From the point of view of interpretation, it is important to know whether the 27-day variation is due to (*a*) screening off of part of the primary radiation, (*b*) the production of additional cosmic ray particles, or (*c*) a modulation of the primary radiation.

Suppose the effect is due to the screening off of part of the primary radiation. In fig. 9 the normal intensity, assuming an isotropic distribution, would be represented by I_1 ; the radiation screened off is represented by the shaded area. Under these circumstances, the mean intensity over a specific 27-day period, I_m, would be higher when the peak-to-peak amplitude was smaller, and vice versa. If on the other hand the effect is due to the production of additional cosmic ray particles, I_2 may be taken as the normal level for isotropic distribution, and the mean intensity, I_m, would be lower when the peak-to-peak amplitude was smaller, and vice versa. In the case of a modulation effect, I_m would represent the normal intensity, and remain constant and quite independent of the peak-to-peak amplitudes.

Fig. 9



Schematic representation of the 27-day variation.

The neutron intensity has a negligible atmospheric temperature effect and is therefore very suitable for investigating these three possibilities. Unfortunately, the day-to-day changes in the neutron intensity were to some extent obscured by the progressive increase in counting rate with time, due to changes in the efficiency of the detecting system. A further correction for this gradual increase in efficiency of the neutron detector

was therefore applied to the barometer-corrected daily intensity totals. A procedure was then adopted which gives a measure of the peak-to-peak amplitudes of separate 28-day periods. This measure, D, is the difference in the mean intensity of a 7-day interval centred around a day of maximum intensity and the mean intensity of a similar interval centred around the day of minimum intensity following the maximum. The procedure for computing D is shown in fig. 10 (*a*). The next step was to calculate the mean intensity of each separate 28-day period, each period being taken as from the 7th day before the maximum. The results are given in table 2.

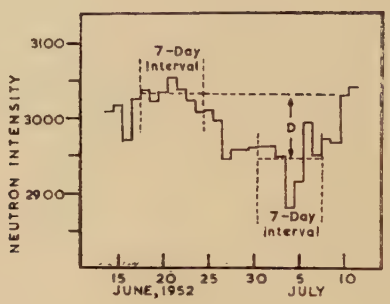
Table 2

No. on graph	Date of maximum	Mean of 7-day interval	Date of minimum	Mean of 7-day interval	D-value	Mean of 28-day interval
	1952 :		1952 :			
1	20 June	3033	3 July	2946	87	2990
2	17 July	3039	3 Aug.	2975	64	3012
3	14 Aug.	3027	31 Aug.	2979	48	3003
4	18 Sep.	3043	27 Sep.	2990	53	3018
5	11 Oct.	3027	16 Oct.	2995	32	3012
6	24 Oct.	3021	1 Nov.	2983	38	3009
7	18 Nov.	3044	27 Nov.	2961	83	3011
8	13 Dec.	3058	28 Dec.	2974	84	3003
	1953 :		1953 :			
9	2 Feb.	3020	17 Feb.	2966	54	2992
10	8 Mar.	3049	16 Mar.	3022	27	3033
11	3 Apr.	3052	15 Apr.	2974	78	3011
12	18 June	3054	7 July	2985	69	3017
13	16 July	3038	28 July	2999	39	3014
14	7 Sep.	3046	25 Sep.	2993	53	3024

These results are shown graphically in fig. 10 (*b*) together with the line of best fit given by the method of least squares; it is assumed that the relationship between the D-value and the mean intensity is linear. The mean intensity for the whole period under consideration was calculated. The D-values were then expressed as percentages of and the mean intensities as percentage deviations from the mean of the whole period. The slope of the line given by the method of least squares is $-0{\cdot}868$. From the large negative slope, it is obvious that for smaller peak-to-peak amplitudes of the 28-day variation, the mean neutron intensity for separate 28-day periods is higher and vice versa, which would seem to indicate that the recurrence tendency is due to a screening off of part of the primary cosmic radiation. This result suggests that the 27-day recurrence tendency is caused by decreases in the intensity which occur at 27-day intervals, and which supports the view of Neher and Forbush (1952) that the 27-day variation

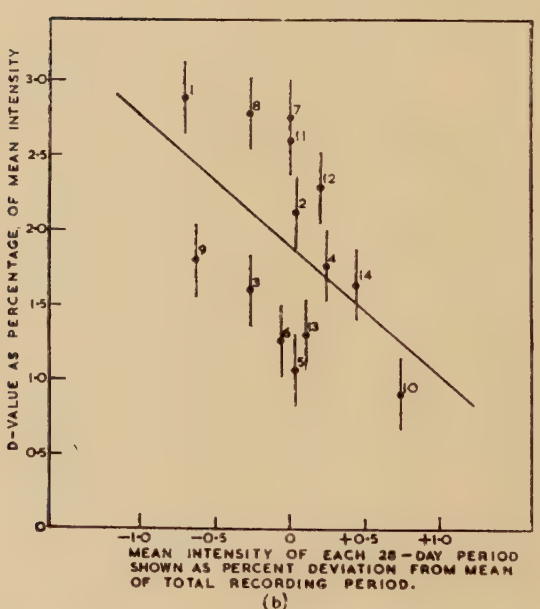
and the large cosmic ray storms are very similar. It therefore seems quite likely that both phenomena can be produced by the same mechanism.

Fig. 10



(a)

Procedure for computing D.



(b)

Value of D plotted as a function of the mean intensity of each 28-day period.

§ 4. Conclusion

The present results confirm the existence of a marked 27-day recurrence tendency for both the neutron and ionizing intensities, and it is found that the amplitude of the 27-day variation is more or less the same for the total cosmic ray flux, the hard and soft components, and approximately one-half that of the neutron intensity. Some evidence is then given

that the 27-day recurrence is caused by decreases in the cosmic ray intensity which occur at 27-day intervals. It can therefore be the same kind of phenomenon as the large cosmic radiation decreases which sometimes occur at the time of intense geomagnetic storms. The 27-day recurrence tendency of the cosmic ray intensity, however, precedes that of magnetic activity by about 5 days, and it is tentatively suggested that the 27-day intensity changes are associated with solar disturbances, which after a delay of a few days, produce geomagnetic field disturbances.

Acknowledgments

We are indebted to Professor P. M. S. Blackett for providing the facilities for this work, and to Dr. H. Elliot for his constant interest and for many valuable discussions. I. J. v. H. wishes to thank the British Council and T. T. the University of Ceylon for maintenance grants.

References

Chree, C.,1913, *Phil. Trans.* A, **212**, 76.
Dawton, D., 1952, *Ph.D. Thesis*, University of Manchester.
Elliot, H., and Dolbear, D. W. N., 1950, *Proc. Phys. Soc.* A, **63**, 137.
Forbush, S. E., 1954, *Journ. of Geophys. Res.*, **59**, 525.
Kane, R. P., 1955, *Phys. Rev.*, **98,** 130.
Neher, H. V., and Forbush, S. E., 1952, *Phys. Rev.*, **87,** 899.
Simpson, J. A., 1954, *Phys. Rev.*, **94,** 426.
Simpson, J. A., Fonger, W., and Treiman, S. B., 1953, *Phys. Rev.*, **90,** 934.

CXXXV. *Ambipolar Thermodiffusion of Electrons and Holes in Semiconductors*

By P. J. Price

I. B. M. Watson Scientific Computing Laboratory, Columbia University, New York City, U.S.A.*

[Received June 6, 1955]

Summary

The purpose of this paper is to bring to notice, and discuss the consequences of, the phenomenon of ambipolar diffusion of electrons and holes down a temperature gradient in a semiconductor. A formula for the magnitude of the ambipolar flux is derived, and is applied to the electronic contribution to the thermal conductivity and to the Nernst Effect. The anomalous thermal conductivity of indium antimonide, reported by Busch and Schneider, cannot be accounted for in terms of the energy transported by this ambipolar flow.

§ 1. Introduction

The purpose of this paper is to bring to notice, and discuss the consequences of, a transport phenomenon in semiconductors which has no counterpart in metals. If a temperature gradient is maintained in a semiconductor then, since the equilibrium concentrations of electrons and holes increase with temperature, there will be a concentration gradient of each in the direction of the temperature gradient. Hence diffusion will cause a flux of electrons, and a flux of holes, *down* the temperature gradient. (This effect is in addition to that due to the Soret Effect† and is in general much greater, as we shall see.) In the absence of an electric field these fluxes cannot be expected to be equal, so there will be a net electric current. If the specimen is left ' open circuit ', an electric field such as to reduce the current to zero will appear. The two fluxes will then be equal, one having been increased and the other decreased by the field, so that the final effect of the temperature gradient is an ambipolar flow of current carriers down the gradient. This flow is of particular interest because it contributes extra terms, which can be large compared with the ' conventional ' ones arising from the electron and hole bands considered separately, to the magnitudes of the electronic thermal conductivity and the Nernst Effect. The latter phenomena are discussed in the following sections. In the present section a formula for

* Communicated by the Author.

† The Soret Effect is the drift due to a temperature gradient, given by the term in D_i^T of eqn. (2).

the magnitude of the ambipolar flux, for conditions as general as is convenient, is derived. The conditions assumed are as follows:

(*a*) The specimen is homogeneous and a single crystal.

(*b*) The crystal has cubic symmetry.

(*c*) The temperature gradient, and its divergence, are not so large as to cause appreciable deviations of the carrier densities and distribution functions from their equilibrium values: the transport equations remain linear.

(*d*) Only a single conduction-band minimum energy (that of the lowest conduction band), and a single valence-band maximum energy (that of the highest valence band) are involved in the transport phenomena. (That is to say the energy distances to the next maximum and to the next minimum are *either* large compared with *or* small compared with kT.) We do not require the conduction (valence) band involved to have a unique energy minimum (maximum) in pseudomomentum space; so that for example germanium and silicon, which have a multiple energy minimum in their lowest conduction bands and a degenerate maximum (two bands touching) in their highest valence bands, are covered by the theory.

(*e*) 'Impurity band' effects (Hung and Gliessmann 1950) are negligible.

(*f*) At the band maxima and minima involved, the energy may be taken to depend quadratically on the pseudomomentum $\mathbf{p}$,

$$\epsilon(\mathbf{p})-\epsilon(\mathbf{p}_0) \simeq A\,|\,\mathbf{p}-\mathbf{p}_0\,|^2, \qquad (1)$$

over an energy range at least of order kT. In (1) A is independent of $|\,\mathbf{p}-\mathbf{p}_0\,|$; but we do not require it to be independent of the direction of $\mathbf{p}-\mathbf{p}_0$, nor that (1) be an analytic function of $\mathbf{p}$ (the 'warped' valence bands of Ge and Si (Dresselhaus *et al.* 1954) are covered).

(*g*) The concentrations of electrons and holes are low enough for Boltzmann statistics to be applied.

Let the number-densities of carriers be n_i and the mobilities, diffusion coefficients and Soret-effect coefficients be μ_i, D_i and D_i^T respectively, where the subscript labels the band: $i=1$ for electrons and $i=2$ for holes. With a temperature gradient grad T and electric field $\mathbf{E}$, the drift velocities $\mathbf{u}_i$ are given by

$$-\mathbf{u}_i=[D_i\,d(\log n_i)/dT+D_i^T]\,\text{grad}\,T\pm\mu_i\mathbf{E}. \qquad (2)$$

If we substitute the Einstein relation

$$D_i=\mu_i kT/e, \qquad (3)$$

and set

$$D_i^T \equiv \gamma_i D_i/T, \qquad (4)$$

in (2), it becomes

$$-\mathbf{u}_i=\mu_i\{(k/e)[T\,d(\log n_i)/dT+\gamma_i]\,\text{grad}\,T\pm\mathbf{E}\}. \qquad (5)$$

For an ' open circuit ' crystal we have equal fluxes:

$$n_1\mathbf{u}_1=n_2\mathbf{u}_2\equiv\mathbf{N}, \text{ say}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The value of $\mathbf{E}$ required, in (5), to satisfy (6) gives the Thomson coefficient. Eliminating $\mathbf{E}$ in (5) then results in

$$\mathbf{N}=-\frac{n_1\mu_1 n_2\mu_2}{n_1\mu_1+n_2\mu_2}\left(\frac{k}{e}\right)\left[T\frac{d(\log n_1 n_2)}{dT}+\gamma_1+\gamma_2\right]\text{grad } T. \quad . \quad (7)$$

Now, on assumptions (*c*), (*d*), (*f*) and (*g*) above, we have

$$n_1 n_2=\text{const. } T^3\exp(-\Delta/kT), \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where Δ is the forbidden energy gap between conduction and valence bands. Then, writing for the conductivities

$$\sigma=\sigma_1+\sigma_2, \qquad \sigma_i=en_i\mu_i, \quad . \quad . \quad . \quad . \quad . \quad (9)$$

(7) becomes (Price 1954)

$$\mathbf{N}=-\frac{\sigma_1\sigma_2}{\sigma}\left(\frac{k}{e^2}\right)\left[\frac{\Delta}{kT}+3+\gamma_1+\gamma_2\right]\text{grad } T. \quad . \quad . \quad (10)$$

We may expect the first term of the square bracket of (10) to be often the dominant one: at room temperature for Ge and Si the first term is about 30 (somewhat smaller for many intermetallics), while for the conventional lattice scattering and spherical energy surfaces γ_1 and γ_2 are each $\frac{1}{2}$. The factor $\sigma_1\sigma_2/\sigma$ of (10) cannot exceed $\sigma/4$. For ' intrinsic ' conditions it is

$$\sigma/[\mu_1/\mu_2+2+\mu_2/\mu_1].$$

The essential origin of this flow process is the fact that conduction electrons and holes are not separately conserved in number: the hotter end of the crystal is a source of pairs and the colder end is a sink of them. An analogous phenomenon is the superfluid ' internal convection ' in liquid helium II, according to the Landau (1941) model, in which there is a flow of excitations (carrying mass, energy and entropy) relative to the centre-of-gravity reference frame. Another analogous phenomenon is the counter currents of a dissociating molecule and its constituents set up by a temperature gradient in a gas (Dirac 1924).

§ 2. Electronic Thermal Conduction

It is obvious that the diffusing pairs must transport the energy required to create them. The energy transported is $\Delta+(\delta_1-\delta_2)kT$, where the factors δ_i have magnitudes of order unity, which will depend on the relaxation processes and band structures (for lattice scattering and spherical energy surfaces they each have the value 2). Hence the total energy transport is

$$\mathbf{Q}=\mathbf{N}[\Delta+(\delta_1+\delta_2)kT]+\mathbf{Q}_1+\mathbf{Q}_2, \quad . \quad . \quad . \quad . \quad (11)$$

where the $\mathbf{Q}_i$ are the standard electronic thermal conduction effects contributed by each band considered separately. Combining (10) and (11), we arrive at an electronic thermal conductivity of form (Price 1954)

$$\left.\begin{aligned} K_{\text{elec}} &= \sigma T(k/e)^2(\Omega+\Omega_0) \\ \Omega(T) &= (\sigma_1\sigma_2/\sigma^2)[(\Delta/kT)^2+\alpha(\Delta/kT)+\beta] \end{aligned}\right\}, \quad . \quad . \quad . \quad (12)$$

where again α and β depend on the detailed relaxation processes, and Ω_0 is the ' conventional ' contribution from the $\mathbf{Q}_i$ of (11). The factor $\Omega+\Omega_0$ replaces that in the conventional Wiedemann–Franz Law (Seitz 1940 a) for metals: clearly in semiconductors it has in general a much larger value. For intrinsic germanium at 300°K, for example, we have

$$\Omega = 0{\cdot}22[29^2+29\alpha+\beta]:$$

this value probably lies between 200 and 300. The anomalously large Wiedemann–Franz ratio for semiconductors does not lead to electronic contributions to the thermal conductivity greater than in metals, because the electrical conductivity is so much smaller for semiconductors. Thus, for Ge and Si at their respective melting points one estimates a contribution to K_{elec} from the term of (12) proportional to Δ^2 of about 0·3 watt units in each case. Since K_{elec} should increase, and the lattice contribution to thermal conductivity decrease, with increasing temperature, this result indicates the possibility of a minimum in the total thermal conductivity, K, fairly near the melting point in each case.

Recently Busch and Schneider (1954) observed a remarkable anomaly, of this kind, in the variation with temperature of the thermal conductivity of indium antimonide. The value of K has a minimum of a third of a watt unit somewhat below 100°C, and has doubled again by a little above 400°C. It can be interpreted as the sum of two terms,

$$K = K_1(T)+K_2(T), \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where K_1 only is appreciable below 0°C. K_1 varies as $1/T$, and may be identified with the lattice contribution. The other term, K_2, varies according to

$$K_2 = 4{\cdot}7 \exp(-0{\cdot}26 \text{ ev}/2kT) \text{ watt units.} \quad . \quad . \quad (14)$$

Busch and Schneider interpret K_2 as electronic in origin, as is very reasonable considering its temperature variation (14). Because of this temperature variation it may be assumed, presumably, that the crystals yielding the published data were intrinsic in the temperature range in question (100°C–500°C). Since electrical conductivity data for the same crystals are not available, we have estimated $K_2/\sigma T$ by using the conductivities published by Tanenbaum and Maita (1953),* in particular for their ' crystal A '. From the data in the range 60° to 364°C, we find

$$K_2/\sigma T \simeq 50(k/e)^2, \text{ experimental.} \quad . \quad . \quad . \quad . \quad (15)$$

* The following numerical analysis is due to Mr. P. B. Linhart.

(The authors' ' crystal B ', evidently more nearly intrinsic, would presumably have given a somewhat larger value for (15) if its conductivity in the same range had been available.) The factor 50 in (15) compares with the value $\Omega_0=2$ (see eqn. (2) of Busch and Schneider) which is obtained by treating each band separately (i.e. not including our diffusion effect) and also assuming lattice-scattering relaxation and spherical energy surfaces.

To compare (15) with (12) we estimate Ω as follows: for $\Delta=0{\cdot}26$ ev and $T\simeq 500^\circ$K, $\Delta/kT\simeq 6$, and hence the first term in the square bracket in (12) is about 36. The other two terms in this case cannot be expected to be negligible in comparison; and in fact for spherical energy surfaces and lattice scattering they are such that the whole bracket comes to 100. However, if we adopt the mobility ratio $\mu_1/\mu_2\simeq 85$ found by Welker (1954), for the intrinsic state the first factor $\sigma_1\sigma_2/\sigma^2$ has a value of about 1/87, so that $\Omega\sim 1$. (The maximum possible value of this first factor is $\frac{1}{4}$, giving $\Omega\sim 25$: we cannot reach the experimental value (15), from (12), even when the energy transport is not reduced by this factor arising from the thermoelectric field equalizing the two fluxes.) Thus it appears that Busch and Schneider's effect cannot be accounted for by the energy transported by the ambipolar thermodiffusion.

It seems desirable that the effect be restudied by measuring both thermal and electrical conductivities, as a function of temperature, *on the same single crystal specimen*, and further observing the effect of strong magnetic fields on both.

§ 3. Nernst Effect

We now consider the effect of a magnetic field $\mathbf{H}$, small enough not to change the mobilities, crossed with the temperature gradient. Since the electrons and holes are drifting in the same direction they will be deflected in *opposite* directions and thus give rise to *reinforcing* transverse currents. The magnitudes of these currents are given each by the product of $(e/c)\mathbf{N}\times\mathbf{H}$ with an appropriate mobility. We now show, subject to assumptions (*b*), (*c*) and (*g*) of § 1, that this is just the Hall mobility for the band. We need only consider a single band in isolation, so drop the subscript i. The Boltzmann equation for the distribution function $f(\mathbf{r}, \mathbf{p})$ is, in our case,

$$-\mathscr{C}f=\frac{\partial\epsilon}{\partial\mathbf{p}}\cdot\frac{\partial f}{\partial\mathbf{r}}+\mathbf{F}\cdot\frac{\partial f}{\partial\mathbf{p}}, \qquad \ldots\ldots (16)$$

where $\mathscr{C}f$ is the ' collision derivative ', the rate of relaxation of f, and $\mathbf{F}$ is the Lorentz force

$$\mathbf{F}=q[\mathbf{E}+(\mathbf{v}\times\mathbf{H})/c] \qquad \ldots\ldots (17)$$

for charge $q=\pm e$ and group velocity $\mathbf{v}(\mathbf{p})=\partial\epsilon/\partial\mathbf{p}$. A Lorentz–Gans solution (Seitz 1940 b) of (16) is obtained by writing f as a series in orders

of the transport ' forces ', $f=f_0+f_1+\ldots$, with the choice of f_0 such that, with the operator $\mathscr{D}$ depending appropriately on the local temperature $T(\mathbf{r})$, $\mathscr{D}f_0=0$:

$$f_0=n(\mathbf{r})A\exp(-\epsilon/kT), \qquad (18)$$

where $A(T)$ is a normalizing factor (proportional to $T^{-3/2}$ on assumption (f) of § 1). We have

$$\frac{\partial f_0}{\partial \mathbf{p}}=-\frac{\partial \epsilon}{\partial \mathbf{p}}\frac{f_0}{kT}, \qquad \frac{\partial f_0}{\partial \mathbf{r}}=f_0\left[\frac{d(\log n)}{d\mathbf{r}}+\left(\frac{d(\log A)}{dT}+\frac{\epsilon}{kT^2}\right)\frac{dT}{d\mathbf{r}}\right]. \qquad (18')$$

The term on the right of (16) involving f_0 and $\mathbf{H}$ vanishes, so that —dropping the terms explicitly proportional to $dT/d\mathbf{r}$,* which we are not interested in here—we find f_1 is a solution of

$$-\mathscr{D}f_1-(q/c)(\mathbf{v}\times\mathbf{H})\,.\,\partial f_1/\partial\mathbf{p}=f_0\mathbf{v}\,.\,[\nabla\log n-\mathbf{E}(q/kT)]. \qquad (19)$$

Thus an electric field $\mathbf{E}$ and a concentration gradient $-n\mathbf{E}(q/kT)$ lead, even in the presence of a magnetic field, to the same disturbance (expressed by $f_1(\mathbf{p})$) of the velocity distribution. For the component of drift velocity parallel to $\nabla\log n-\mathbf{E}(q/kT)$, this result merely verifies the Einstein relation (3) : for the perpendicular component it shows that the ' Hall ' mobilities for drift currents due to an electric field and to concentration diffusion (our ambipolar effect is a combination of the two) are the same.

If B is the Nernst coefficient, defined by

$$\mathbf{E}_N=B\nabla T\times\mathbf{H}, \qquad (20)$$

as measured by the transverse field $\mathbf{E}_N$ for virtually infinite surface recombination rate (Landauer and Swanson 1953), then from the foregoing results we find

$$B=(\mu_1{}^H+\mu_2{}^H)(\sigma_1\sigma_2/\sigma^2)(k/ec)[\Delta/kT+3+\gamma_1+\gamma_2]+B_0 \qquad (21)$$

where $\mu_i{}^H$ are the Hall mobilities and B_0 is the ' genuine ' Nernst coefficient contributed by each band taken separately. To indicate the magnitude of the effect we estimate B from (21) for intrinsic germanium at 300°K, by dropping B_0 and setting $\gamma_1+\gamma_2\sim 1$. (Strictly speaking, supposing $\gamma_i=\frac{1}{2}$ is consistent only with also supposing $\mu_i{}^H=3\pi\mu_i/8$, which is not true ; but the γ_i make a small contribution to the square bracket of (21), so we give them this somewhat arbitrary value while using experimental values for the $\mu_i{}^H$.) Substituting $\Delta=0{\cdot}75$ ev and the values for μ_i and $\mu_i{}^H$ found by Morin (1954), we obtain

$$B\simeq 4{\cdot}3\times 10^{-8} \text{ volts/gauss degree, theoretical estimate.} \qquad (22)$$

* These are responsible for the Soret term (the term in D^T) of eqn. (2).

This is a considerably larger value than, to my knowledge, has been found by measurement on any substance. For comparison, a table of selected published experimental values follows:

Substance	Temperature, and B in volts/gauss degree	Source
Bi Te Pb Ir Sn Si	$-119°$C, 9×10^{-9}; $62°$C, $4{\cdot}6\times10^{-10}$ $33°$C, $3{\cdot}6\times10^{-9}$; $204°$C, $2{\cdot}7\times10^{-9}$ $57°$C, -5×10^{-14} $18°$C, -5×10^{-14} $57°$C, ? -4×10^{-14} $47{\cdot}4°$C, $3{\cdot}3\times10^{-10}$ (Gottstein 1914) $50°$C, $1{\cdot}1\times10^{-10}$ and $2{\cdot}0\times10^{-10}$ (Buckley 1914)	*International Critical Tables* 1929, **6**, 420
PbTe PbSe	$50°$C, $1{\cdot}4\times10^{-9}$ (maximum value) $420°$C, 4×10^{-10} (maximum value)	Putley 1955

For intrinsic silicon, (21) would predict a value for B of order 10^{-8} volts/gauss degree. However, appreciable doping could reduce the value considerably: if we drop the B_0 term from (21), and consider various dopings, at fixed temperature, not so large as to change the mobilities, then the constancy of n_1n_2 leads to the rule*

$$B\sigma^2=\text{constant}=(B\sigma^2)_{\text{intrinsic}}. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

In Buckley's paper he mentions that his Si sample had several per cent of impurities (mostly Fe, some Al, and smaller fractions of other elements). This amount is, of course, far beyond the range of validity of (23), because of the effect of impurities on mobility and on band structure and crystal perfection, apart from the fact that B_0 will no longer be the negligible term in (21). It would evidently be of interest to make Nernst Effect measurements on Ge and Si crystals of present-day purities.

Putley (1955) quotes a formula which reduces to

$$B=\frac{3\pi}{8}(\mu_1+\mu_2)\frac{\sigma_1\sigma_2}{\sigma^2}\left(\frac{k}{ec}\right)\left[\frac{\Delta}{kT}+4\right]-\frac{3\pi}{16}\left(\frac{k}{ec}\right)\frac{n_1{\mu_1}^2+n_2{\mu_2}^2}{n_1\mu_1+n_2\mu_2} \quad . \quad . \quad (24)$$

in the present notation and units. It is evident by comparison with (21) that

(*a*) the first term of (24) corresponds to the first term of (21) and the second term of (24) to the second term (B_0) of (21); and

(*b*) (24) has been derived by assuming spherical energy surfaces and lattice scattering.

Thus (21) is a generalization of (24); but (24) gives an explicit expression for B_0 for a special case. In the conditions of Putley's observations on the intermetallic semiconductors PbTe and PbSe, the two terms of (24) were

* This result might be used as a test to determine when B_0 is negligible compared with B.

of comparable magnitude. Because they are of opposite sign, the Nernst coefficient as a function of temperature passed through a zero in each case.

Since in practice semiconductor crystals even of highest purity become non-intrinsic at a low enough temperature, the first (thermodiffusion) term of (21) will have a maximum as a function of temperature, falling off to low values at both ends of the scale.* For a reasonably pure crystal of a semiconductor with a large energy gap and not too unequal mobilities (such as Ge and Si), however, there should be a range of temperature where we may both neglect B_0 and approximate this first term by

$$B \simeq (\mu_1{}^H+\mu_2{}^H)(\sigma_1\sigma_2/\sigma^2)(\phi/cT) \quad . \quad . \quad . \quad . \quad (21')$$

where $\Delta \equiv e\phi$. We consider now what information could be obtained by determining B for a fixed temperature and different dopings, in conditions where (21′) is a good approximation. It is convenient to define a ' Nernst mobility '

$$\mu^N \equiv B(cT/\phi)=(\mu_1{}^H+\mu_2{}^H)\sigma_1\sigma_2/\sigma^2, \quad . \quad . \quad . \quad (25)$$

just as the Hall mobility is defined by

$$\mu^H \equiv \sigma R=(\mu_1{}^H\sigma_1-\mu_2{}^H\sigma_2)/\sigma. \quad . \quad . \quad . \quad . \quad (26)$$

Using the fact that $\sigma_1+\sigma_2=\sigma$, we may eliminate the σ's from (25) and (26). Then

$$\mu^N=(\mu_1{}^H-\mu^H)(\mu_2{}^H+\mu^H)/(\mu_1{}^H+\mu_2{}^H). \quad . \quad . \quad . \quad (27)$$

Over the doping range for which the band mobilities are constant, the plot of μ^N against μ^H should thus lie on a parabola. The vertex of the parabola (maximum of μ^N) should be for the same doping as that for which the conductivity is a minimum, and have the coordinates

$$\mu^H=\tfrac{1}{2}(\mu_1{}^H-\mu_2{}^H), \qquad \mu^N=\tfrac{1}{4}(\mu_1{}^H+\mu_2{}^H). \quad . \quad . \quad . \quad . \quad (28)$$

The determination of the point (28) thus would give the values of $\mu_1{}^H$ and $\mu_2{}^H$ separately, ϕ being known. This is the same information as is obtainable from the Dunlap ellipse (Dunlap 1950)—though in the present method it is obtained from a single portion of the doping curve, that around the resistivity maximum. However, if there were difficulty in getting absolute values of B, or uncertainty over the value of ϕ, the present method would determine only $\mu_1{}^H-\mu_2{}^H$: this is just what can be determined from the neighbourhood of the resistivity maximum by the Dunlap ellipse. Furthermore (21′) is only an approximation, with a limited area of usefulness.

---

* There will in consequence be a change in sign of B (to negative at the lowest temperatures) where extrinsic conditions set in. There can also be a second change of sign (to negative at the highest temperatures), if (21) becomes negative (B_0 dominant) when we set $n_1=n_2$ and $\Delta/kT=0$. (In Putley's formula (24)—that is, for spherical energy surfaces and lattice scattering—the condition for this is that *either* $\mu_1>7{\cdot}87\mu_2$ *or* $\mu_2>7{\cdot}87\mu_1$: this condition is satisfied for InSb.) It might be, however, that sufficiently heavy doping could depress the first term of (21) far enough for there to be no changes of sign, the value of B remaining negative at all temperatures.

There is an analogy between the phenomenon, contributing to the Nernst Effect, discussed above and the phenomenon underlying the photomagnetoelectric effect (Moss, Pincherle and Woodward 1953), in which electrons and holes generated by light falling on a semiconductor surface diffuse inward together and are deflected by a magnetic field parallel to the surface, so that they maintain an electric field perpendicular to the magnetic field and parallel to the surface.

Acknowledgments

I am indebted to Mr. P. B. Linhart for numerical analysis of experimental data and for other assistance in the preparation of this paper; to Dr. Frank Herman for information on the band structure of indium antimonide; to Dr. Conyers Herring for informing me of Busch and Schneider's results; to Dr. R. G. Breckenridge for a discussion on various experimental values for InSb; to Professor I. Prigogine for bringing to my notice the gas transport phenomenon, referred to at the end of § 1, studied by Nernst, by Dirac, and by himself and his co-workers; to Dr. G. L. Tucker for helpful discussions and information; and to Dr. G. R. Gunther-Mohr and Dr. S. H. Koenig of this laboratory and to the referee for helpful comments on the manuscript.

References

Buckley, O. E., 1914, *Phys. Rev.*, **4,** 482. Buckley does not specify a sign convention, but it seems likely that he used the convention of the original 1887 Nernst paper.

Busch, G., and Schneider, M., 1954, Amsterdam Conference Proceedings, *Physica*, **20,** 1084.

Dirac, P. A. M., 1924, *Proc. Camb. Phil. Soc.*, **22,** 132. See also Nernst, W., 1904. *Boltzmann Festschrift*. p. 504; and Lafleur, S., Prigogine, I., and Waelbroeck, F., 1955, *Physica*, to be published.

Dresselhaus, G., Kip, A. F., and Kittel, C., 1954, *Phys. Rev.*, **95,** 568.

Dunlap, W. C., 1950, *Phys. Rev.*, **79,** 286.

Gottstein, G., 1914, *Ann. der Phys.*, **43,** 1079.

Hung, C. S., and Gliessmann, J. R., 1950, *Phys. Rev.*, **79,** 726.

Landau, L., 1941, *J. Phys. U.S.S.R.*, **5,** 71.

Landauer, R., and Swanson, J., 1953, *Phys. Rev.*, **91,** 555.

Morin, F. J., 1954, *Phys. Rev.*, **93,** 62.

Moss, T. S., Pincherle, L., and Woodward, A. M., 1953, *Proc. Phys. Soc.* B, **66,** 743.

Price, P. J., 1954, *Phys. Rev.*, **95,** 596. Also Fröhlich, H., and Kittel, C., 1954, Amsterdam Conference Proceedings, *Physica*, **20,** 1086.

Putley, E. H., 1955, *Proc. Phys. Soc.* B, **68,** 35. See also Johnson, V. A., and Lark-Horovitz, K., 1948, *Phys. Rev.*, **73,** 1257.

Seitz, F., 1940 a, *Modern Theory of Solids* (New York: McGraw-Hill), § 32, eqn. (21); 1940 b, *Ibid.*, § 31.

Tanenbaum, M., and Maita, J. P., 1953, *Phys. Rev.*, **91,** 1009.

Welker, H., 1954, Amsterdam Conference Proceedings, *Physica*, **20,** 893. Breckenridge, R. G., *et al.* (Amsterdam Conference Proceedings, *Physica*, **20,** 1073), however, have proposed the value 29.

CXXXVI. CORRESPONDENCE

Observations on Etched Crystals of Barium Titanate

By D. S. Campbell
The Plessey Co. Ltd., Caswell Research Laboratories, Towcester, Northants.*

[Received August 19, 1955]

Observations of the etching effect of cold concentrated hydrochloric acid on barium titanate single crystals have been reported by Hooton and Merz (1955). We have obtained similar results using hydrochloric acid and phosphoric acid on crystals direct from the melt, with no polarizing treatment. Figures 1 and 2 (figs. 1–4, Plate 28) show the effect of etching a crystal containing two wedges of optically active material ('a' axis normal to the surface).

Figure 1 shows the crystal examined with crossed nicols and monochromatic light under which conditions the interference fringes produced by the wedges are clearly visible. Figure 2 shows the surface of the crystal after etching in hydrochloric acid for 1 minute. The positive ends of the 'c' domain areas etch most heavily, the 'a' areas less heavily and the negative ends of the 'c' areas least of all. It is to be noted that fig. 2 shows the 'c' domains running parallel to the wedges. That the 'c' domains run vertically through the crystal is shown in figs. 3 and 4, where the correspondence between the top and bottom surfaces is very good.

Fig. 5

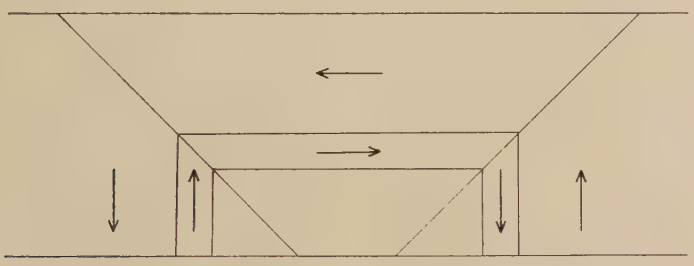


Diagram of coupling through a wedge.

Examination of a wedge shows coupling through the wedge as illustrated in fig. 5. Dark areas, unetched, at one side of the wedge, correspond to light areas, etched, on the other. Figures 6 and 7 (figs. 6–9, Plate 29) show the type of micrograph obtained.

* Communicated by H. F. Kay.

Observation of several wedges has shown that coupling does not hold after a certain distance (approximately 0·07 mm) indicating some breakdown in the head to tail domain arrangement over relatively large distances.

Thicker crystals have shown similar results. Figure 8 shows the top and fig. 9 the bottom surface of a crystal 0·16 mm thick and it can be seen that coupling does exist but is very poor, the vast majority of the domains present at the top surface not appearing at the bottom.

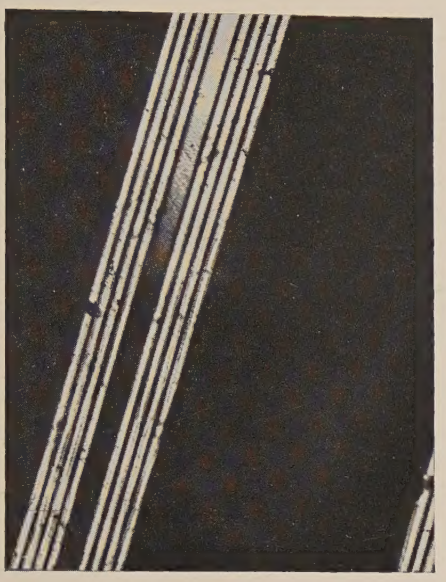
The etching technique has proved useful for the examination of barium titanate crystals. Failure of head to tail coupling after a certain minimum distance is observed, so that thin crystals should respond more easily than thick ones to applied fields. An electron microscopy study of the etched surface is being undertaken and will be reported elsewhere.

The author would like to thank the Plessey Co. for permission to publish this note.

Reference

Hooton, J. A., and Merz, W. J., 1955, *Phys. Rev.*, **98**, 409.

Fig. 1



Interference pattern produced by ' a ' wedges. ×100.

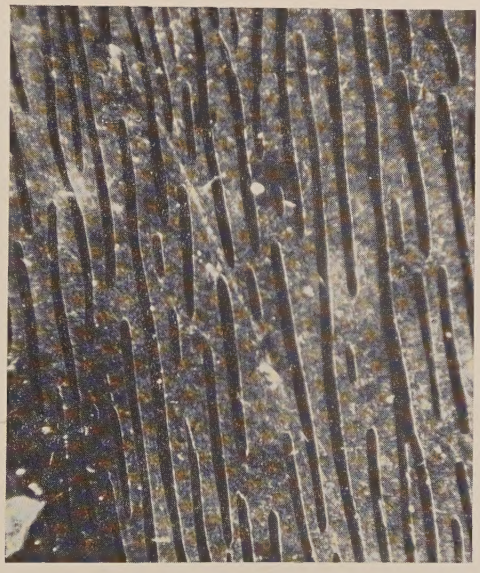
Fig. 2

Effect of etching in conc. HCl for 1 min. ×100.

Fig. 3

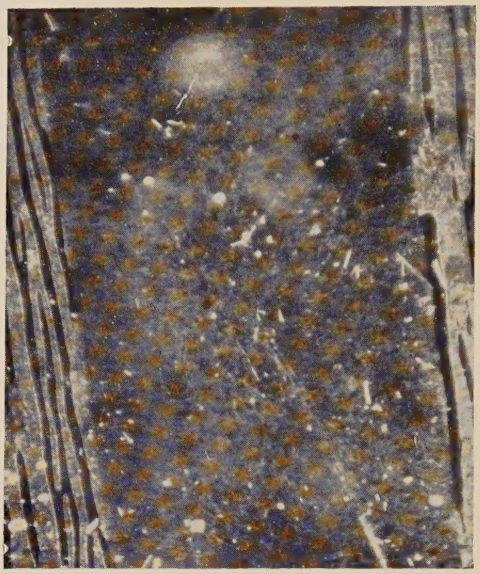
Top surface of ' c ' region crystal 0·048 mm thick. Ultropak. ×530.

Fig. 4



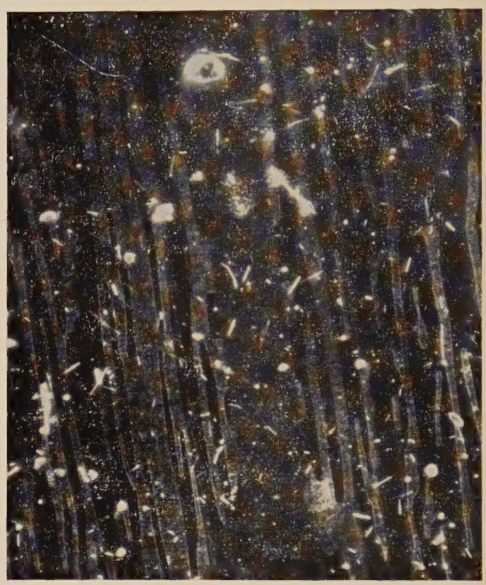
Bottom surface of ' c ' region crystal 0·048 mm thick. Ultropak. ×350.

Fig. 6



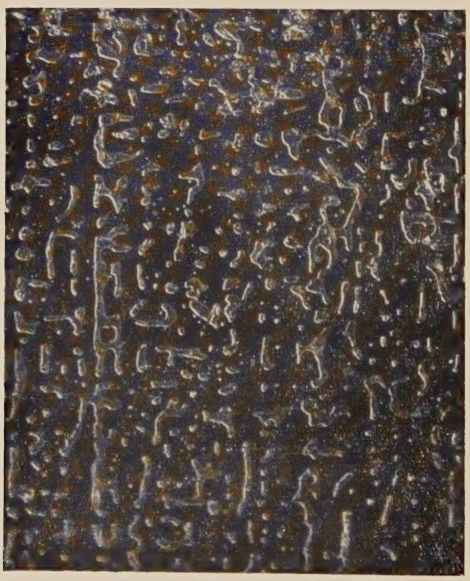
Top surface of wedge area.
Ultropak. ×350.

Fig. 7



Bottom surface of wedge area showing reflection. Ultropak. ×350.

Fig. 8



Top surface of ' c ' region crystal 0·16 mm thick. Ultropak. ×200.

Fig. 9

Bottom surface of ' c ' region crystal 0·16 mm thick. Ultropak. ×200.

CXXXVII. REVIEWS OF BOOKS

Solubilization and Related Phenomena. By M. E. L. McBain and E. Hutchinson. [Pp. xv+259.] (Academic Books Ltd.) Price 56*s*.

Solubilization is an awkward word coined to express the increased take-up of a wide variety of substances in certain colloidal solutions. The widespread technological importance of the phenomenon is familiar from the enormous growth of the detergent industry ; less generally familiar technical applications of solubilization are found in practices such as emulsion polymerization, the solubilization of dyestuffs for treating textiles and food, and the solubilization of proteins in certain processes for the manufacture of fibres.

This monograph aims to reduce a wealth of empirical knowledge and application to some kind of fundamental principles. After a brief survey of the historical background and of the properties of colloidal electrolytes some 95 pages are devoted to a very useful factual survey of experiments on solubilization.

Some kind of aggregate or micelle formation plays an essential part in the process. But discussions of the mechanism of solubilization as presented in this book make it clear that experimental evidence on the precise structure of such micelles still is very scanty. The present review of the experimental findings and of the current controversies of interpretation will be found valuable both by specialists in this field, and by those who are interested in the progress of physico-chemical theory generally. A. R. Ubbelohde.

Optical Properties of Thin Solid Films. By O. S. Heavens. [Pp. 261.] (London: Butterworths Scientific Publications, 1955.) Price 35*s*.

This book, by Dr. O. S. Heavens, dealing with the optical properties of thin solid films is a valuable and timely contribution to a subject which, within recent years, has grown to be of extreme importance both scientifically and technologically. Written by an expert the book covers (1) the formation of thin films, (2) the structure of thin films, (3) thin film optics, (4) the measurement of thin-film thickness and optical constants, and (5) practical applications of thin films in optics. The experimental production of solid thin films is admirably described by an author who himself has clearly had much practical experience in the production of such films. Of particular value is the discussion of the electron-microscopic studies of the structures of thin films. Thin-film optics, which is a formidable subject, is treated sympathetically and with the minimum of mathematical analysis needed for the study of such a complex subject. This section is unavoidably the heaviest in the book, yet at the same time will probably prove to be the most valuable in that it will save the reader rom the difficult task of reading many papers, mostly in foreign languages, and at times obscure. It is particularly in this chapter, that dealing with film optics, that the author renders a considerable service to those who are working in this field, especially to the practically minded man. The various experimental methods available for measuring thin-film thicknesses are discussed in detail. These include mechanical methods, electrical methods, optical methods and interferometric methods. The use of radioactive tracers is briefly indicated. The last chapter, dealing with practical applications, covers a very wide field such as that of anti-reflection systems, interference films, interference filters, multilayers, etc. and although consisting of a bare 50 pages is in fact a summary of a very considerable number of substantial publications.

The book is practically free from blemish. The reviewer has only noted in two minor instances incorrect allocation of priority for some idea or experiment. This is of little consequence.

Production is excellent and the cost not excessive as far as modern prices go. There is no doubt that this will become a most valuable source of reference to all workers in the field of thin-film optics. The book is hardly to be recommended for the undergraduate, and was not intended for readers at this level. It was intended for the research worker, and as such, we welcome this as a valuable contribution to a growing and difficult subject. S. T.

Quantum Theory of Solids. By R. E. Peierls. [Pp. vi+229.] (Oxford: Clarendon Press, 1955.) Price 30*s*.

The chief danger to a student of solid state physics is that he will acquire a glib facility for manipulating various phenomenological concepts (e.g. phonon, hole, collision, dislocation, relaxation time) without really understanding their limitations. Professor Peierls begins with the atomic lattice and deduces its behaviour—thermal and electrical conduction, cohesion, magnetism, interaction with light, superconductivity—directly from first principles. The book exhibits a rare rigour, not of pedantic mathematical formalism but in its faith to the physical model. The omission of numbers, graphs, detailed experimental results and complicated analytical formulae leaves the argument unobscured, whilst real fundamental difficulties are emphasized. The lucid exposition of early work not previously available in English is especially valuable, although there are places (e.g. in discussing electron–electron interaction) where the theory is not quite up to date. Students will find it simple and clear, making an admirable introduction to the basic theory; all who work in the field will benefit from the criticism of many fallacious over-simplifications which are widely current. J. M. Z.

High-Energy Accelerators. By M. S. Livingston. (New York and London: Interscience Publ. Inc., 1954.) [Pp. viii+157.] Price 26*s*.

There could hardly be an author more suitable for this kind of review than Professor Livingston, who has been constructing high energy accelerators from their very beginning, and as a general common-sense introduction to the subject the book could not be better. Most of it is concerned with the actual physics of the various accelerating systems; only the first tenth of the book discusses their applications and the reasons for the drive towards ever higher energies. After the second chapter, in which the general physical principles are explained (relativistic motion, oscillations about stable orbits etc.), the next four chapters are concerned with the four main types of high energy machines: electron synchrotron, synchrocyclotron, linear accelerator and proton synchrotron. The last chapter describes the alternating-gradient focusing ('strong focusing') which has made possible the latest push towards energies above 10^{10} electron volts. Mathematical formulae are given where needed but their derivation is only briefly indicated. Points of engineering are mentioned but not described in detail, and there is only brief mention of the problems of shielding.

The construction of high energy machines has become an extremely complex and highly specialized subject which nobody could expect to learn from a book, but as a quick survey or as an introduction to the more specialized reports, this little book will serve admirably. O. R. F.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]